WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

(11) International Publication Number:

WO 98/07864

26 February 1998 (26.02.98)

(21) Internati nal Application Number:

(43) International Publication Date:

A1

PCT/GB97/02273

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4 9625996.5

23 August 1996 (23.08.96)

GB 13 December 1996 (13.12.96) GB

Research), Porton Down, Salisbury, Wiltshire SP4 0JG (74) Agent: SCHLICH, George, William; Mathys & Squire, 100

Conrad, Padraig [GB/GB]; Microbiological Research

Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology &

(60) Parent Application or Grant

(63) Related by Continuation

US Filed on 08/782,893 (CIP)

27 December 1996 (27.12.96)

(71) Applicants (for all designated States except I/S): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN. Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑŪ	Australia	GA	Gabon	1. V	Latvia	S7.	Swaziland
AZ.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	18	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Vict Nam
CG	Солдо	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation	1	
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sti Lanka	· SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

WO 98/07864 PCT/GB97/02273

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the H_c domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the H_N domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The H_N domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the $H_{\rm c}$, amino acid residues 449-871 for the $H_{\rm N}$ and residues 1-448 for the LC. Digestion with trypsin effectively degrades the $H_{\rm c}$ domain of the BoNT/A to generate a non-toxic fragment designated $LH_{\rm N}$,

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolat d LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

- (A) clostridial neurotoxin light chain:
- -a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.
 - (B) clostridial neurotoxin heavy chain H_N domain:
- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H_c domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a cl stridial t xin H_C domain.

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polyp ptid of the inv ntion may be obtain d by expression of a r combinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain pr ferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the inv ntion described in an xample below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodim nt of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (ii) a C-terminal extension to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, link d for example by disulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated $H_{\rm C}$ of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated $H_{\rm C}$ of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C₁, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then d termines the targ t for a polypeptide - immunoglobulin compl x. Alt rnatively, the

WO 98/07864 - 8 - PCT/GB97/02273

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endop ptidas activity of the composition will, of course, also be d termined by the amount of the diviative that is present.

While it is known to treat naturally produced clostridial toxin to remove the $H_{\rm C}$ domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described abov.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polyp ptide or a fusion prot in and comprises nucleotides encoding residu s 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH₇₂₈/B (SEQ ID NO: 20).

• The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recov r d from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_c domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nom nclature,

LH₄₂₃/A = SEQ ID NO: 2, containing the entire L-chain and 423 amino acids of the H-chain of botulinum neurotoxin type A;

₂LH₄₂₃/A = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

₂L_{/2}H₄₂₃/A = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain;

²L_{FXa/2}H₄₂₃/A = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

 $_2$ L_{FXə/2}H₄₂₃/A-IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention d signated LH₄₂₃/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated ${}_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/4H423/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated L_{FXa/3}H₄₂₃/A illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated L_{FXe/3}H₄₂₃/A-IGF-1, a fusion protein; the IGF-1 sequence begins at position G₈₈₂;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Ω_{882} ;
- Fig.9 sh ws th C-terminal portion of the amino acid sequence of an

embodiment of the invention designated $L_{FXa/3}H_{423}/A-ZZ$, a fusion protein; th C-terminal ZZ sequence begins at position A_{890} immediately after a genenase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11

polypeptides of the invention; Fig. 10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an lg binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12

shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig. 13

shows Western blot analysis of recombinant LH₁₀₇/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH₁₀₇/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH₁₀₇/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH₄₂₃/A

2	LH ₄₂₃ /A
3.	DNA coding for $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$), of which an N-terminal portion is shown in Fig. 4.
4	₂₃ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
5	DNA coding for ₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-
	terminal portion is shown in Fig.4
6	₂ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
7	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L _{/4} H ₄₂₃ /A
10	L _{/4} H ₄₂₃ /A
11	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A
12	L _{FXa} / ₃ H ₄₂₃ /A
13	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
14	L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
16	L _{FX3} / ₃ H ₄₂₃ /A-CtxA14
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	L _{FXa/3} H ₄₂₃ /A-ZZ
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$)
26	LH ₄₂₃ /A (Q ₂ E,N ₂₈ K,A ₂₇ Y)
27	DNA c ding for LH ₄₁₇ /B wherein th first 274 bases are

WO 98/07864

28

modified to have an *E. coli* codon bias

DNA coding for LH₄₁₇/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_c) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the LH_{423}/A coding sequence was PCR amplified from total chromosomal DNA from Clostridium botulinum and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polym rase amplified DNA (bas s 914-1138 and 1976-2616) and the r mainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced bγ addition of isopropylthio-β-D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH423/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH₄₂₃/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH₄₂₃/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH_{423}/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A .

Two further variants of the gene sequence that have been expressed and the corresponding products purified are $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a *Eco* 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the *Eco* 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, L_{/4}H₄₂₃/A (SEQ ID NO: 10), which c ntains an additional four amino acids between amino acids 448 and 449 of LH₄₂₃/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa/3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L_{/4}H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) $L_{\text{FXe/3}}H_{423}/A\text{-CtxA14}$ (SEQ ID NO: 16) , in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) $L_{\text{FXe/3}}H_{423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A (Q_2 E, N_{26} K, A_{27} Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 μ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, 1.8 mM KH $_2$ PO $_4$, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

₂LH₄₂₃/A is more stable in the presenc of trypsin and more active in the in vitro

peptide cleavage assay than is $_{23}LH_{423}/A$. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH₇₂₈/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polyp ptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native $\text{LH}_\text{N}/\text{A}$.

A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Anoth r variant's quenc equival nt to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-₂LH₄₂₃/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 μ g of GST- $_2$ LH $_{423}$ /A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH $_N$ /A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na $_2$ HPO $_4$ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 μ g (5X10-6 g) to 5 picograms (5X10-12 g). Aliquots of 0.5 ml were injected into mice intrap ritoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- $GST_{-2}LH_{423}/A$ antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST- $_2$ LH $_{423}$ /A antiserum.

Botulinum	Toxin/mouse

Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	-	4	4	4	4	4	4
4		4	4	4	4	4	4 .

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

Botulinum Toxin/mouse

Survivors On Day	0.5µg	0.0 05µ g	0.0005µg	0.5ng	0.005ng	5pg	Control
. 1	0	0	0	0	o	2	4
2	-	•	•	-	-	0	4
3 .	-	-	-	•		-	4
4	•	-	•	•	•	•	4

Example 4

Expression of recombinant LH₁₀₇/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH_{107}/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in *E. coli* BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH₄₂₃/A.

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
 - (B) STREET: 14 Kensington Square
 - (C) CITY: London
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): W8 5HH
 - (A) NAME: FOSTER; Keith Alan
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) .NAME: QUINN; Conrad Padraig
 - (B) STREET: Centre For Applied Microbiology And Research,
 Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: SHONE; Clifford Charles
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E), COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

								_			-					
ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC	AAC Asn	TAT Tyr	Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn	GGT	48
GTT Val	GAC Asp	ATT	GCC Ala 20	lyr	ATC Ile	AAA Lys	ATT	CCA Pro 25	AAC Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Pne	AAG Lys	ATT	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GAÇ Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT . Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720

	.			24	_	111.	L AS	II AI	25	r 1y 0	r G1	.u Me	t Se	er G	ly 55	Leu	768
, 01			26	0	NG GA .u Gl	ם מפו	ı MI	26	5 PR	e GI	y G1	y Hi	s As 27	р А. 0	la 1	Lys	816
TT Ph	T AT e Il	C GA e As 27	, P	C TI	G CA	G GAO n Glu	AA(As: 28(I GT	G TT u Ph	C CG	r Cr g Le	G TA u Ty 28	r Ty	C TA	AC A	VAC Asn	864
AA(Lys	G TT s Ph 29	,	A GA s As	T AT	T GC/ e Ala	A AGT Ser 295	T 111	CTC Leu	AA(1 Ası	AAC Lys	GC' 6 Ala 30	a Ly:	G TC s Se	C AT	T C	TG 'al	912
GGT Gly 305		C AC	T GC r Al	T TC a Se	A TTA r Leu 310	. 0111	TAI	Met	AAA Lys	AAI Asn 315	va.	T TT:	r AA ≥ Ly:	A GA 5 Gl	u L	AA ys 20	960
TAI Tyr	CTC	C CT	A TC u Se	T GA r Gl: 32	A GAT u Asp 5	ACA Thr	TCT Ser	GGA Gly	Lys 330	Phe	TCC Ser	GTA Val	A GAT	1 AA 2 Ly 33	s L	TA eu	1008
AAA Lys	Phe	GA' As	T AAG D Ly: 340	s Let	A TAC	AAA Lys	ATG Met	TTA Leu 345	Thr	GAG Glu	ATT	TAC Tyr	ACA Thr	Gl	G G	AT sp	1056
AAT Asn	Phe	GT: Va: 35		TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Tyr	TTO	3 Ai	AT sn	1104
TTT Phe	GAT Asp 370		A GCC S Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AA7 Asr	T	AC 'r	1152
ACA Thr 385	ATA Ile	TAT	GAI Asp	GGA Gly	Phe	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AA As	n	1200
		U _1	U	405	ACA Thr	GIU	116	ASR	410	Met	Asn	Phe	Thr	Lys 415	Le	u	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GIY	TTG Leu	TTT Phe	GIU	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AG. Ar	A 9	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	THE.	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AA Ly	3	1344
	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC I Ile 1 455	AAA (Lys '	GTT .	AAT Asn	Asn '	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TT:	.	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT 1 Phe 1	ACT I	AAT (Asn)	Asp .	CTA I Leu I 475	TAA neA	AAA Lys	GGA Gly	GAA Glu	GAZ Glu 480	l	1440
ATT I	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	AAT ; Asn ;	ATA (BAA (Blu /	rra v	GCA (Ala (190	GAA (Glu (SAA Slu	AAT : Asn :	Ile .	AGT Ser 495	TTA Leu	.	1488
GAT 1	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT 1	rat i Cyr I	en 1	ACC Thr I	Phe 1	AAT 1 Asn E	TT (Asp A	AAT (Asn (GAA Glu	CCT Pro	•	1536

GAA AAT ATT TCA ATA GAA AAT CTT TCA AGT GAC ATT ATA GGC CAA TTA Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520 GAA CTT ATG CCT AAT ATA GAA AGA TTT CCT AAT GGA AAA AAG TAT GAG 1632 Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu TTA GAT AAA TAT ACT ATG TTC CAT TAT CTT CGT GCT CAA GAA TTT GAA 1680 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 555 CAT GGT AAA TCT AGG ATT GCT TTA ACA AAT TCT GTT AAC GAA GCA TTA 1728 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 TTA AAT CCT AGT CGT GTT TAT ACA TTT TTT TCT TCA GAC TAT GTA AAG 1776 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys AAA GTT AAT AAA GCT ACG GAG GCA GCT ATG TTT TTA GGC TGG GTA GAA 1824 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA GTA AGT ACT ACG 1872 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT ATA GGA CCT GCT 1920 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 630 635 1968 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA CCA GAG ATT GCA 2016 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT ATT GCG AAT AAG 2064 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT AAA AGA AAT GAA 2112 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT TGG TTA GCA AAG 2160 Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG AAA GAA GCT TTA 2208 Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC TAT CAG TAT AAT 2256 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 2304 Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT ATG ATT AAT ATA 2352 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 780

- 33 -

785	-,-		TTG Leu	ASII	790	. Cys	Ser	vai	ser	795	Leu	Met	Asn	Ser	Met 800	2400
ATC Ile	CCT Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	Tyr	ATT Ile 870	AAG Lys	TAA *									2616

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 520

PCT/GB97/02273

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys
675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys 705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 860

Thr Phe Thr Glu Tyr Ile Lys * 865

(2) INFORMATION FOR SEQ ID NO: 3:

WO 98/07864 PCT/GB97/02273 - 36 -

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS (B) LOCATION: 1.. 2685

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

								_		-						
GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	48
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	144
AAG Lys	TAC Tyr 50	GGC Gly	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	AAG Lys 60	ATT Ile	CAT His	AAC Asn	AAA Lys	192
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	240
GAC Asp	TTG Leu	AAC Asn	CCG Pro	CCG Pro 85	CCG Pro	GAA Glu	GCA Ala	AAG Lys	CAG Gln 90	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr 95	TAC Tyr	288
GAT Asp	TCA Ser	ACC Thr	TAT Tyr 100	CTG Leu	AGC Ser	ACA Thr	GAC Asp	AAC Asn 105	GAG Glu	AAG Lys	GAT Asp	AAC Asn	TAC Tyr 110	CTG Leu	AAG Lys	336
GGA Gly	GTG Val	ACC Thr 115	AAA Lys	TTA Leu	TTC Phe	GAG Glu	CGT Arg 120	ATT Ile	TAT Tyr	TCC Ser	ACT Thr	GAC Asp 125	CTG Leu	GGC Gly	CGT Arg	384
ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	ATC Ile	GTC Val 135	CGC Arg	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser	432
ACC Thr 145	ATT Ile	GAC Asp	ACG Thr	GAG Glu	TTG Leu 150	AAG Lys	GTT Val	ATT Ile	GAC Asp	ACT Thr 155	AAC Asn	TGC Cys	ATT Ile	AAC Asn	GTG Val 160	480
					AGC Ser											528
					GAC Asp											576
CAC His	GAA Glu	GTG Val 195	TTG Leu	AAC Asn	CTG Leu	ACG Thr	CGT Arg 200	AAC Asn	GGT Gly	TAC Tyr	GGC Gly	TCT Ser 205	ACT Thr	CAG Gln	TAC Tyr	624

AT'	CG' Are	a bu	C AG e Se:	C CCI	A GAC	TTC Phe 215	Ini	TTO Phe	C GG: e Gly	T TT(y Phe	GA(Gl) 22(7 GJ	G AG u Se	C CT r Le	G GAG u Glu	672
GT: Val 225	LAS	T AC	C AA(r Ası	CCC Pro	CTC Leu 230	ı Leu	GGT	GC/ Ala	A GGO a Gly	235	: Phe	GC.	A AC a Th	T GA	T CCA P Pro 240	720
GCC Ala	GT(G ACC	C CTO	GCA Ala 245	His	GAG Glu	CTG Leu	ATC Ile	CAC His 250	3 Ala	GG1 Gly	CA'	CG' Arg	T CT(3 Let 25!	G TAT u Tyr 5	768
GIŞ	, 116	: Ala	260	ASI	Pro	Asn	Arg	Val 265	. Phe	: Lys	Val	Asr	270	(Asi	GCC Ala	816
ıyı	Tyr	275	Met	Ser	GIY	Leu	280	Val	Ser	Phe	Glu	Glu 285	Let	Arc	ACG Thr	864
Pne	290	GIA	, HIS	Asp	Ala	Lys 295	Phe	Ile	Asp	Ser	Leu 300	Gln	Glu	Asn	GAG Glu	912.
305	Arg	Leu	Tyr	ıyr	310	Asn	Lys	Phe	Lys	Asp 315	Ile	Ala	Ser	Thr	CTG Leu 320	960
Asn	rys	Ala	Lys	Ser 325	Ile	Val	Gly	Thr	Thr 330	Ala	Ser	Leu	Gln	Tyr 335		1008
rys	Asn	Val	9he 340	Lys	GIu	Lys	Tyr	Leu 345	Leu	Ser	Glu	Asp	Thr 350	Ser	-	1056
Lys	Pne	Ser 355	Val	Asp	Lys	TTA Leu	Lys 360	Phe	Asp	Lys	Leu	Tyr 365	Lys	Met	Leu	1104
Thr	370	Ile	Tyr	Thr	Glu	GAT Asp 375	Asn	Phe	Val	Lys	Phe 380	Phe	Lys	Val	Leu	1152
385	Arg	Lys	Thr	Tyr	190	AAT Asn	Phe .	Asp	Lys	Ala 395	Val	Phe	Lys	Ile	Asn 400	1200
116	Val	Pro	Lys	Val 405	Asn	TAC . Tyr	Thr	Ile	Tyr 410	Asp	Gly	Phe	Asn	Leu 415	Arg	1248
Asn	Thr	Asn	Leu 420	Ala	Ala .	AAC ' Asn	Phe .	Asn 425	Gly	Gln .	Asn	Thr	Glu 430	Ile	Asn	1296
ASN	Met	Asn 435	Phe	Thr	Lys :		Lys 1 440	Asn	Phe	Thr	Gly	Leu 445	Phe	Glu	Phe	1344
Tyr	Lys 450	Leu	Leu	Cys	Val i	AGA (Arg (455	Gly :	Ile	Ile	Thr	Ser 460	Lys	Thr	Lys	Ser	1392
TTA Leu 465	GAT Asp	AAA Lys	GGA Gly	Tyr .	AAT A Asn 1 470	AAG (Lys)	SCA :	ITA . Leu .	Asn .	GAT 1 Asp 1 475	TTA ' Leu '	TGT Cys	ATC Ile	Lys	GTT Val 480	1440

701			, wah	485	File	; PILE	: Sei	Pro	490	Glu	ı Ası) Ası	n Phe	3 Th:	-	1488
YO	, 100	. ngu	500	Cly	GIU	GIU	116	505	ser	Asp	Thr	· Asr	1 Ile 510	e Glu	A GCA 1 Ala	1536
nac	· GIU	515	. vem	110	361	neu	520	. rea	ıııe	GIn	Gln	525	Туг	Let	ACC Thr	1584
riic	530	FILE	rsp	ASII	GIU	5 35	GIU	Asn	ı Ile	Ser	11e 540	Glu	Asn	Leu	TCA Ser	1632
545	Asp	116	116	GIY	550	Leu	GIU	Leu	Met	Pro 555	Asn	Ile	Glu	Arg	Phe	1680
PIO	ASII	GIY	Lys	565	lyr	GIU	ren	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	TAT Tyr	1728
CTT Leu	CGT	GCT Ala	CAA Gln 580	GAA Glu	TTT Phe	GAA Glu	CAT His	GGT Gly 585	AAA Lys	TCT Ser	AGG Arg	ATT Ile	GCT Ala 590	TTA Leu	ACA Thr	1776
ASII	SEI	5 9 5	AAC Asn	GIU	Ата	Leu	600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe	1824
FIIC	610	261	GAC Asp	lyr	vai	615	rys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala	1872
625		Leu	GGC Gly	irp	630	GIU	GIn	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640	1920
Int	Set	GIU	GTA Val	645	inr	Tnr	qsA	Lys	11e 650	Ala	Asp	Ile	Thr	Ile 655	Ile	1968
116	PIO	TYE	ATA Ile 660	GIY	PIO	Ala	Leu	Asn 665	Ile	Gly	Asn	Met	Leu 670	Tyr	Lys	2016
Asp	Asp	675	GTA Val	GIÀ	ATS	ren	680	Phe	Ser	Gly	Ala	Val 685	Ile	Leu	Leu	2064
GAA Glu	TTT Phe 690	ATA Ile	CCA Pro	GAG Glu	TIE	GCA Ala 695	ATA Ile	CCT Pro	GTA Val	TTA Leu	GGT Gly 700	ACT Thr	TTT Phe	GCA Ala	CTT Leu	2112
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	ALA	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720	2160
GCT Ala	TTA Leu	AGT Ser	AAA . Lys .	AGA Arg 725	TAA neA	GAA Glu	AAA Lys	TGG Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile	2208
GTA Val	ACA Thr	ASI	TGG Trp 740	TTA Leu	GCA . Ala	AAG Lys	Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	qeA	CTA Leu 750	ATA Ile	AGA Arg	2256

- 39 -

		75	5	. 010		- 260	760)	ı GII	ı Ala	a GIU	1 Ala 765	a Th	r Ly	G GCT s Ala	2304
	770)	, -	J 2.	,-	775	GII	TY	Ini	GIU	780	Glu	Ly:	s Ası	AAT Asn	2352
785			• • • • • • • • • • • • • • • • • • • •		790	nap	Deu	261	ser	795	Leu	Asn	Glı	Ser	ATA Ile 800	2400
	-,-			805	ASII	116	ASII	Lys	910	Leu	Asn	Gln	Cys	Ser 815		2448
	-,-		ATG Met 820		JCI	MEL	116	825	Tyr	GIA	Val	Lys	Arg	Leu	Glu	2496
		835	GCT Ala	501	Deu	пуъ	840	MIG	ren	Leu	Lys	Tyr 845	Ile	Tyr	Asp	2544
	850	01,	ACT Thr	neu	116	855	GIN	vai	Asp	Arg	Leu 8 60	Lys	Asp	Lys	Val	2592
865			CTT Leu	361	870	veħ	116	PIO	Pne	875	Leu	Ser	Lys	Tyr	GTA Val 880	2640
GAT Asp	AAT Asn	CAA Gln	AGA Arg	TTA Leu 885	TTA Leu	TCT . Ser '	ACA Thr	Pne	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	AAG Lys	TAA * 895		2685

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu 1 5 10 15

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30 -

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro
35 40 45

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 200 Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 265 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 280 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 465 470 475 480

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 485 490 495

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 500 505 510

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 530 535

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 545 550 555 560

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 575 575

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 580 585 590

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 595 605

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 610 615 620

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 625 630 635 640

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 645 650 655

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
660 665 670

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 675 680 685

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 690 695 700

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 705 715 720

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
725 730 735

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 740 745 750

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
755 760 765

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 770 780

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile 790 795 800

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 805 810 815 Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 835 840 845

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 850 855

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 865 870 875 880

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 895 890 890

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2622

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA Gly 1	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT Tyr	AAG Lys	GAC Asp	CCT Pro 15	GTA Val	48
AAC Asn	GGT Gly	GTT Val	GAC Asp 20	ATT Ile	GCC Ala	TAC Tyr	ATC Ile	AAA Lys 25	ATT Ile	CCA Pro	AAG Lys	TAC Tyr	GGC Gly 30	CAG Gln	ATG Met	96
CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
AGC Ser	ACA Thr	GAC Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	288
TTC Phe	GAG Glu	CGT Arg	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu	384
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432

ACC TAC AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val 11e 11e Gly Pro Ser Ala 155 GAC ATT ATC CAG TTT GAC TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC ASD 11e 11e Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn 165 CTG ACC CGT AAC GGT TAC GGC TCT ACT CAG TAC AAT CGT TCC AGC CCA Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 11e Arg Phe Ser Pro 180 GAC TTC ACC TTC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG ASP Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 210 CAC GAG CTG ATC CAC GCC GCT CAT CGT CTG TAT GGC ATT ACC His Glu Leu 11e His Ala Gly His Arg Leu Tyr Gly 11e Ala 11e Asn 225 CCG AAC CGC GTG TC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATT ACC His Glu Leu 11e His Ala Gly His Arg Leu Tyr Gly 11e Ala 11e Asn 225 CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATT ACC Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACT TTAC GAG ATT ACC GTY ASN ARG VAl Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 246 GGT TTA GAA GTA AGC TTC GAG GAA ACC AAC GAC TTC GGT GGC CAT GAT GTY Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 GCC AAG TTT ATC GAC AGC TTC CAG GAG AAC GAG TTC GGT GGC CAT GAT GTY Asn Lys Phe Lie Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 275 GCC AAG TTT ATC GAC AGC TTC CAG GAG AAC GAG TTC CGT CTG TAC TAC Ala Lys Phe 11e Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 290 ATT GTG GGT ACC ACT CT TCA TTA CAG TAT ATA AAA ATT TTC GAT GAT Ala CAAC AAG TTT AAA GAT ATT GAA ATT ACA TAT AAT AAA TTT TTA AAA GAT ATT AGA GAT AAA ATT TTC GAT GAT ALE VAL Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 305 ATT GTG GGT ACC ACT CTT TTA AAA ATA TTT TAA AAA ATT TTC GAT AAT 1104 GAG GAT AAT TTT GAT AAA GTT TTT AAA GAT AAA TTT TAC GAT AAT 1104 GAG GAT AAT TTT GAT AAA GTT TTT AAA GAT AAA ATT TAA CAA ATT TTY 316 GAG GAT AAT TTT GAT															•							
CCG AAC CGC GAC ACC GCC GGT CAC CGC CGC CAC CAC CAC CAC CAC ACC CGA CCC GAC CTC GAC CGC CAC CGC CG		14	45	•		J			150		u no	iii De	u va	1.	1e 1	.ie	Gly	Pro	S€	r Al 16	a 0	480
GAC TTC AGG TCC GGT TTC GAG GAG AGC CTG GAG GTT GAT ACC AAC CCG ASP Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 205 205 205 205 205 205 205 205 205		G# As	AC sp	ATI Ile	IA T	ic ci		••••	GAG Glu	TG: Cy:	C AA s Ly	G AG	r Pr	ie G.	GC C	AC (GAA Glu	GTC Val	. Le	u As	C n	528
200 Set led Git val Asp Thr Asn Pro 201 205 CTG TTG GGT GCA GGC AAG TTC GCA ACT GAT CCA GCG GTG ACC CTG GCA Leu Leu Giy Ala Gly Lys Phe Ala Thr Asp Pro Ala val Thr Leu Ala 210 215 CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn 225 230 CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACC TTC GGT GGC CAT GAT Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 265 GCG AAC TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT TAC TAC Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 275 TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA GTG AAC AAG GCC TAC TAC AAC AAC AAG TTT AAA GAT ATT CAC AGT ACA GTG AAC AAC GAG TTC CGT TAC 280 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA 305 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA 305 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AATT GTT TTA AAA 306 GAA AAA TAT CTC CTA TTA GAG AGT ACA CTG GAA AAA TTT TCG GTA GAT Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 325 AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAG GAT AAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATA ACA TTT 340 GAG GAT AAT TTT GAT AAG GTT TAT AAG ATA ATT ACA AGA AAA ACA TAT 340 TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA CTA ACA ATA TTT TY 345 AAT TAC ACA ATA TAT GAT GAG GAT TTT AAT TAA GAT AAT A						18	30	,	-7-	92)	, 3e.	18	5	.n 13	/r I	⊥e ≯	Arg	Phe 190	Se	r Pro	0	576
CAC GAG CTG ATC CAC GCC GGT CAT CGT CTG TAT GGC ATT GCG ATT AAC HIS Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn 220 CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 CGG TTA GAG GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 CGC AAG TTA ACC ACC TTC GAG GAA CTG CGC ACG TTC CGT CTG TAC TAC Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 280 CGC AAG TTT ATC GAC AGC TTG CAG GAG AAC AAC GAC ATC GAT ATC TYR ASN Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 290 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATC AAA AAT GTT TTT AAA 11e Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 305 GAG AAA TAT CTC CTA TCT GAA GAT ACA CTC GAA AAA TTT TAC ACA CAAA TAT AAA TTT GAT AAG TTA TAC AAA ATG TTA TAC ACA CAAA TAT TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAAA TTT AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAAA TAT TTT GAT AAA GTT TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAAA TAT TTT GAT AAA GTT TTA AAA ATG TTA ACA AAA ACA TAT CAAC AAT ATT TTT GAT AAA GTT TTA AAA ATG TTA ACA AAA ACA TAT CAAC AAT TTT GAT AAA GTTA TAC AAA ATG TTA ACA AAA ACA TAT CAAC AAT TTT GAT AAA GTTA TAC AAA ATG TTA ACA AAA ACA TAT CAAC AATA TTT GAT AAA GTTA TAC AAA ATG TTA ACA AAA ATG TTA CACA CAC GAG GAT ATT TTT GAT AAA TTT TAC AAA ATG TTA ACA AAA ATG TTA TAC CAAA AAA TTT GAT AAA GTTA TAC AAA ATG TTA ACA AAA ACA TAT CAAC AATA TTT GAT AAA CCC GTA TTT AAA GTA ACT AACA TAT CAAC AATA TTT GAT AAA CCC GTA TTT AAA ATG ATA ATA ATA GTA ACA TAT CAAC AATA TAT GAT GAT GAT TTA AAA ATT AAA ATA TATA GTA ACA TAT CAAC AATA TAT GAT GAT GAT TTA ACA GAA ATT ATA ATA ATA ATA ATA ATA ATA AT			•		19	5	_	- ,		010	200	0	r ne	n G1	u V	31 A	lsp 105	Thr	Ası	n Pro	•	624
CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT 245 CCG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT 245 CGG AAC CGC GTG TTC AAG GTT AAC ACC AAC GCC TAC TAC GAG ATG AGT 245 GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GIly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 CGC AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT CTG TAC TAC Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 280 TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC 275 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA GTA CTC GLU Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 330 AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA ATG TTT TCG GTA GAT GLU Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 330 AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 345 GAG GAT AAT TTT GTT AAG TTT TTA AAA GTA CTT AAC AGA AAA ACA TAT GLU Asp Asn Phe Val Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr 355 AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CAC Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 365 GAG GAT AAT TTT GAT AAG GTT TTT AAA GTA CTT AAC AGA AAA ACA TAT GLU Asp Asn Phe Val Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr 365 AAT TAC ACA ATA TAT GAT GGA GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA ACA TAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG ATA TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA ATA TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG ATA TTT AAG ATA TAT AAT TAA AAT TTA AAA TTA AAT TTA AAA TTA AAT TAT AAT TAA AAT TTA AAT TAA AAT TTA AAT TAA AAT TTA AAT TAA AAT TTA AAT TAA AAT ACA AAT TTA ACA AAT TTA AAT TAAT AAT		CT Le			GG G1	T GC Y Al	A G	GC ly	AAG Lys	FILE	. WT	A AC	r GA r As	T CC p Pr	O A.	La V	TG al	ACC Thr	CTO	GCA Ala	A	672
GGT TTA GAA GTA AGC TTC GAG GAA CTG CGC ACG TTC GGT GGC CAT GAT GIV Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 260 GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT TAC TAC ALL Lys Phe 1le Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 285 TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC 777 Asn Lys Phe Lys Asp 1le Ala Ser Thr Leu Asn Lys Ala Lys Ser 290 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA 11e Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 320 GAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT GLU Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 335 AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA CTG Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu 1le Tyr Thr 355 GAG GAT AAT TTT GAT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT GLU Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr 3355 TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA ATA TAC AGA AAA ACA TAT TTT GAT AAA GCC ATT TAC AGA ATA ATA GTA CTA AGA ATA TTT GAT AAA GCC ATT TAC AGA ATA ATA TTT GAT AAA GCC ATT TAC AGA ATA ATA TTT GAT AAA GCC ATT TAC AGA ATA TTT GAT AAA GCC ATT TAC AGA ATA ATA TTT GAT AAA GCC ATT TAC AGA ATA ATA TTT GAT AAA GCC ATA TTT AAC AGA ATA TTT GAT AAA GCC ATA TTT AACA ATA ATA ATA ATA ATA ATA		22	5				- 		230	dry	nis	, WI	, re	u 1y 23	r G1 5	y I	le /	Ala	Ile	240		720
GCG AAG TTT ATC GAC AGC TTG CAG GAG AAC GAG TTC CGT TAC TAC ALL Lys Phe lle Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr 285 ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT TTT GGT AGT ACT GLU Lys Tyr Leu Leu Ser Glu Asp The Ser Gly Lys Phe Ser Val Asp 325 AAA TTA AAA TTT GAT AAG TTT TAA GAT ATG AAA ATG TTA ACA GAG ATT TAC ACA ASP Lys Phe Asp Lys Leu Tyr Tyr 326 GAG GAT AAT TTT GAT AAA GCC GTA TTT AAA GTA ATT AAT ATA AAT ATG AAA TTT ACT ALA AAC TTT AAT GCA AAC TTT AAT GAT AAA TTT AAT GAT AAA TTT AAT GAT AAA TTT TA CAC AAT TTR CAC AAT TAT GAT AAA TTT TA CAC AAT TTR AAT T						,	2	45	-y-s	401	ASI	Int	25()	а Ту	T T	yr (Glu	Met 255	Ser		768
TAC AAC AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC 290 295 295 295 295 300 Lys Ala Lys Ser 300 295 300 Lys Ala Lys Ser 300 295 300 295 300 295 300 Lys Ala Lys Ser 300 295 300 29		,	_		-	26			FIIC	GIU	GIU	265	Arc	J Thi	r Ph	e G	ly o	31y 170	His	Asp		816
ATT GTG GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA 960 lie Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 320 gAG AAA TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT ASp Thr Ser Gly Lys Phe Ser Val Asp 335 and 1008 lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 lie Tyr Thr 340 los for a sp Thr Ser Gly Lys Phe Ser Val Asp 335 and 1056 lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 los for a sp Thr Ser Gly Asp Thr Ser Gly Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 los for a sp Thr Ser Gly Lys Phe Asp Lys Phe Lys Val Leu Asn Arg Lys Thr Tyr 360 los for a sp Thr Ser Gly Lys Phe Lys Ile Asn Ile Val Pro Lys Val Lys Val Lys Phe Lys Ile Asn Ile Val Pro Lys Val Asp Asp Tyr Thr Ile Tyr Asp Gly Phe Asp Leu Arg Asp Thr Asp Cly Asp Gly Phe Asp Leu Arg Asp Thr Asp Leu And Asp Asp Thr Asp Gly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr Asp Cly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr Asp Phe Thr Asp Cly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr Tyr Asp Cly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr Tyr Asp Cly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr Tyr Asp Cly Gln Asp Thr Glu Ile Asp Asp Met Asp Phe Thr					275			. بر.		Deu	280	GIU	ASD	GIT	ı Ph	e Ar 28	rg I 15	eu	Tyr	Tyr		864
GAG GAT AAT TIT GTT AAG TIT TIT AAA GTA CTT AAC AGA AAA ACA TAT GAT AAA GCC GTA TTT AAG ATA AAT ATA GAA AAT TTT GAT AAA GCC GTA TTT AAT TAC ASA ATT TAC AAA TTT GAT AAA GCC GTA TTT AAT TAC AAA TTA AAT AAT AAT AAT A		TAC			AAG Lys	TTT	Ly	vs A	. و	T T C	GCA Ala	AGT Ser	ACA Thr	CTG Leu	Ası	n Ly	AG G 's A	CT . la :	AAG Lys	TCC Ser		912
AAA TTA AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 GAG GAT AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr 365 TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 370 AAT TAC ACA ATA TAT GAT GGA GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 385 GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Met Asn Phe Thr				G 1	GGT Gly	ACC	AC Th		Ta 3	CA Ser	TTA Leu	CAG Gln	TAT Tyr	Met	Lys	A AA B As	AT G	TT :	TTT Phe	Lys		960
GAG GAT AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT Glu Asp Asn Phe Val Lys. Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr 365 TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 370 AAT TAC ACA ATA TAT GAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr			-1	•	-,-	200	32	5	er c	J.u	Asp	Inr	330	GIY	Lys	Ph	e S	er (/al	Asp		1008
TTG AAT TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 370 AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr	1	AAA Lys	TI	A A	AAA Lys		GA'	T A	AG I YB I	TA eu	IYL	nys	ATG Met	TTA Leu	ACA Thr	GA:	u I	le 1	TAC Tyr	ACA Thr		1056
AAT TAC ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr	Ċ	GAG Glu	GA As	٠,	1211	TTT Phe	GT: Va	r Ai l L	AG T ys. P	ne .	rne	AAA Lys	GTA Val	CTT Leu	AAC Asn	Arg	g L}	A A 's T	CA hr	TAT Tyr		1104
385 390 395 Asn Thr Asn Leu Ala 400 GCA AAC TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr	I	TTG Leu		-	TT he	GAT Asp	AAA Lys	A GO	La V	αT 1	Phe	AAG Lys	ATA Ile	AAT Asn	Ile	GT/ Val	A CC	T A	AG ys	GTA Val		1152
405 405 406 406 406 416 Asn Asn Met Asn Phe Thr	•		TA(C A	CA hr	ATA Ile	TA1		D G	GA 1	Phe	AAT Asn	TTA Leu	Arg	AAT Asn	ACA Thr	A AA : As	T T	eu 2	Ala		1200
	G A	CA la	AA(Ası	T P	TT he	AAT Asn	GTA	GI	A Ai	AT A	Chr (. IL	TTE	AAT Asn	AAT Asn	ATG	AA As	n P	he :	ACT Thr		1248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	CTA	TGT Cys		1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC Tyr		1344
AAT Asn	AAG Lys 450	GCA Alá	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu		1392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480		1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	TAA Asn	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile		1488
AGT Ser	TTA Leu	GAT Asp	TTA Leu 500	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 505	TTA Leu	ACC Thr	TTT Phe	AAT Asn	TTT Phe 510	GAT Asp	AAT Asn		1536
GAA Glu	CCT Pro	GAA Glu 515	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 520	AAT Asn	CTT Leu	TCA Ser	AGT Ser	GAC Asp 525	ATT Ile	ATA Ile	GGC Gly		1584
CAA Gln	TTA Leu 530	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 535	ATA Ile	GAA Glu	AGA Arg	TTT Phe	CCT Pro 540	AAT Asn	GGA Gly	AAA Lys	AA G Lys		1632
TAT Tyr 545	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 550	ACT Thr	ATG Met	TTC Phe	CAT His	TAT Tyr 555	CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 560		1680
TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 565	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 570	ACA Thr	AAT Asn	TCT Ser	GTT Val	AAC Asn 575	GAA Glu	:	1728
GCA Ala	TTA Leu	TTA Leu	AAT Asn 580	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 585	ACA Thr	TTT Phe	TTT Phe	TCT Ser	TCA Ser 590	GAC Asp	TAT Tyr	:	1776
GTA Val	AAG Lys	AAA Lys 595	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 600	GAG Glu	GCA Ala	GCT Ala	ATG Met	TTT Phe 605	TTA Leu	GGC Gly	TGG Trp	•	1824
GTA Val	GAA Glu 610	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 615	TTT Phe	ACC Thr	GAT Asp	GAA Glu	ACT Thr 620	AGC Ser	GAA Glu	GTA Val	AGT Ser	:	1872
ACT Thr 625	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 630	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 635	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 640	:	1920
CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 645	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 650	AÄA Lys	GAT Asp	GAT Asp	TTT Phe	GTA Val 655	GGT Gly	:	1968
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu	:	2016
ATT Ile	GCA Ala	ATA Ile 675	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	TTT Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala	:	2064

	69	0.				69	5	T 11(e As	P As:	70	a Le	u Se	r Ly	A AGA s Arg	2111
70	5	,	- 	P	71	0	- 1y.	r by:	s ry:	715	e Va. 5	l Th:	r Ası	n Tr	G TTA P Leu 720	2160
	- - ,			72	5	. 11	: ASI	p Let	730	Arg	; Lys	5 Lys	Met	Ty:	_	2208
			74	0		. 010	, WIG	745	rys	Ala	lle	: Ile	750	Туз	CAG Gln	2256
-,-		75	5		. 010	. GIU	760	ь	Asn	Asn	Ile	Asn 765	Phe	Asn	ATT	2304
	770			- UC1	. uya	775	Asn	GIU	ser	He	Asn 780	Lys	Ala	Met	ATT	2352
785			. 2,5		790		GIN	Cys	ser	795	Ser	Tyr	Leu	Met	Asn 800	2400
				805	GIY	GTT Val	гåа	Arg	810	Glu	Asp	Phe	Asp	Ala 815	Ser	2448
	-,-		820	Dea	Deu	AAG Lys	TYE	825	ıyr	Asp	Asn	Arg	Gly 830	Thr	Leu	2496
	O1,	835	V Q 1	rap	ALG	TTA Leu	840	Asp	rys	Val	Asn	Asn 845	Thr	Leu	Ser	2544
	850			• • • • • • • • • • • • • • • • • • • •	GIII	CTT Leu 855	SEL	гуs	Tyr	Val .	GAT Asp 860	AAT Asn	CAA Gln	AGA Arg	TTA Leu	2592
TTA Leu 865	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 870	TAT Tyr	ATT Ile	AAG ' Lys	TAA *							2622

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro 35 40 45

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly 135 Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn 170 Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp. Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala 215 His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 250 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 390 395

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 405 410 415

Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys 420 425 430

Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr 435

Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu 450 460

Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 470 475 480

Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 485 490 495

Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn 500 505 510

Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly 515 520 525

Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 530 535 540

Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu 550 555 560

Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 565 570 575

Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr 580 585 590

Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp 595 600 605

Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser 610 620

Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 635 630

Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly 645 650

Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 660 665 670

Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 675 680 685

Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg 690 695 700

Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu 705 710 715 720

Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu
725 730 735

Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln
740 745 750

-	45	₹ _

										- 40	•					
Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765		Asn	Ile	
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile	
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 800	
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser	
Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	11e 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu	
		835			Arg		840					845				
Thr	Asp 85 0	Ile	Pro	Phe	Gln	Le u 855	Ser	Lys	Tyr	Val	Asp 86 0	Asn	Gln	Arg	Leu	
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*							
(2)	INF	ORMA!	CION	FOR	SEQ	ID N	10:	7:								
		(1 (1) MOI) FEA	3) TY C) ST D) TO LECUI ATURE	(PE: TRANI OPOL LE T' E: AME/I	i: 26 nucl DEDNE DGY: (PE:	leic ESS: line DNA	acidou) ar (ger	d ole								
	(xi)	SEC)UENC	E DE	ESCRI	PTIC	ON: S	SEQ 1	D NO): 7:						
ATG Met 1	CCA Pro	TTT Phe	GTT Val	AAT Asn 5	AAA Lys	CAA Gln	TTT Phe	AAT Asn	TAT Tyr 10	AAA Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT Gly	4.8
GTT Val	GAT Asp	ATT Ile	GCT Ala 20	TAT Tyr	ATA Ile	AAA Lys	ATT Ile	CCA Pro 25	AAT Asn	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	CAA Gln	CCA Pro	96
GTA Val	AAA Lys	GCT Ala 35	TTT Phe	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	ATA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg	. 144
GAT Asp	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA Pro	GAA Glu	192
GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA Pro	GTT Val 70	TCA Ser	TAT Tyr	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80	240
GAT Asp	AAT Asn	GAA Glu	AAA Lys	GAT Asp 85	AAT Asn	TAT Tyr	TTA Leu	AAG Lys	GGA Gly 90	GTT Val	ACA Thr	AAA Lys	TTA Leu	TTT Phe 95	GAG Glu	288

***	,		10	0	I AS	T CT p Le	u GI	10	15 ME	ec Le	u Le	eu Th	r Se 1:	er I LO	le	Val	336
AGC Arg	G GG G Gl	A AT y Il 11	C	A TT o Ph	T TG e Tr	G GG	r GG y Gl 12	y se	T AC	A AT	A GA e As	T AC p Th	r G	AA T lu L	TA eu	AAA Lys	384
GT1 Val	110 130	- na	T AC p Th	T AA r As	T TG n Cy	T ATT s Ile 135	- AS	T GT n Va	G AT 1 I1	A CA e Gl	A CC n Pr 14	O As	T GG p Gl	T A	GT er	TAT Tyr	432
AGA Arg 145	361	A GA	A GA	A CT	T AA' u As: 15	T CTA n Leu 0	GT/	A AT.	A AT e Il	A GG e Gl	y Pr	C TC.	A GC	T G	AT Sp	ATT Ile 160	480
ATA Ile	CAC Glr	TTT Phe	Γ GAJ e Glu	A TG: 1 Cys 165	s пå:	A AGC s Ser	TTT Phe	GG Gly	A CA' Y Hi: 17	s Glu	A GT 1 Va	T TTO	G AA 1 As	T CT n Le	≥ u	ACG Thr	528
CGA Arg	AA1 Asn	GI Gly	TAT TY1	. Gi	TC? Sei	C ACT	CAP Gln	TAC Ty: 185	2 176	r AGA	TT:	T AGO	C CC.	o As	T	TTT Phe	576
ACA Thr	TTT Phe	GGT Gly 195	FILE	GAC Glu	GAG Glu	TCA Ser	CTI Leu 200	GIU	A GTT	GAT L Asp	ACA Thi	A AAT Asn 205	Pro	CT Le	T 1	ITA Leu	624
GGT Gly	GCA Ala 210	GTA	Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA	GCA Ala	GTA Val	ACA Thr 220	Leu	GC/ Ala	A CA A Hi	T (s (GAA Glu	672
CTT Leu 225	ATA Ile	CAT	GCT Ala	GGA Gly	CAT His 230	AGA Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT	AA1 Asr	CC.	o A	AT Asn 40	720
AGG Arg	GTT Val	TTT Phe	AAA Lys	GTA Val 245	AAT Asn	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	Tyr	GAA Glu	ATG Met	AGT Ser	GG(G1; 25;	y L	TA eu	768
GAA Glu	GTA Val	AGC Ser	TTT Phe 260	GAG Glu	GAA Glu	CTT Leu	AGA Arg	ACA Thr 265	TTT Phe	GGG Gly	GGA Gly	CAT His	GAT Asp 270	GC/ Ala	A A	AG ys	816
TTT Phe	ATA Ile	GAT Asp 275	AGT Ser	TTA Leu	CAG Gln	GAA Glu	AAC Asn 280	GAA Glu	TTT Phe	CGT Arg	CTA Leu	TAT Tyr 285	TAT Tyr	TAT	A	AT sn	864
AAG :	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	G'	TA al	912
GGT A Gly 3 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GA G	L	AA ys 20	960
TAT (CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	Le	ra eu	1008
AAA 1 Lys E	TT Phe	rsp	AAG Lys 340	TTA Leu	TAC Tyr	AAA ; Lys i	met .	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	G# As	AT Sp	1056
AAT I	-11	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	TAS A	STA (Val : 360	CTT Leu	AAC Asn	AGA . Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AA As	AT on	1104

TTT Phe	GAT Asp 370	гÀг	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	TATA	GTA Val	CCI Pro	Lys	GTA Val	AA1 Asr	TAC Tyr	1152
385	ITE	Tyr	Asp	GIY	390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	AAC Asn 400	1200
. Pne	Asn	GIA	GIn	405	Inr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	Val	AGA Arg	1296
Gly	Ile	11e 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	•	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
Ile	Thr	Ser	Asp	Thr 485	AAT Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu	1488
Asp	Leu	Ile	Gln 500	Gln	TAT Tyr	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro	1536
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	TAA Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT Tyr	GAG Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 10	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	Ąsp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920

_,				64	5	c Det	1 1Y1	L Ly	650	P As _l	p Phe	e Val	. Gly	/ Al 65		1
			66	0		- 116	: Let	669	u GI(ı Phe	≥ Ile	Pro	670	ı Il	T GCA e Ala	1
	- 11.	67	5	u	y 111.	. FILE	680	Let	ı val	. Ser	Tyr	Ile 685	Ala	Ası	T AAG n Lys	
GT Va	T CTA l Lev 690		C GT r Va	T CA l Gl	A ACI	A ATA F Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	Lys	AGA Arg	AA: Asi	GAA 1 Glu	2112
AA Ly: 70:		G GA	T GA	G GT u Va	C TAT 1 Tyr 710	. Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	TAA 1 naa l	AC.	A CAG	G AT: n Ile 729	= veh	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	Leu	2208
GAA Glu	TAA A	CA.	A GCA n Ala 740	2 310	A GCA 1 Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
921		755	5	. GIU	GIU	AAA Lys	760	Asn	ile	Asn	Phe	Asn 765	Ile	Asp	Asp	2304
	770	50.	. Dys	Dea	. ASII	GAG Glu 775	Ser	TTE	Asn	Lys	780	Met	Ile	Asn	Ile	2352
785	2,5	- ***	neu.	Vall	790	TGC Cys	ser	val	ser	795	Leu	Met i	Asn	Ser	Met 800	2400
ATC Ile	CCT Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GII	GAT Asp 810	TTT Phe	GAT Asp	GCT / Ala S	Ser :	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA '	TAL 1	GAT Asp 825	AAT . Asn .	AGA Arg	GGA / Gly !	Thr I	TTA I	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT A	AAA (Lys 1 340	GTT . Val :	AAT / Asn /	AAT :	Thr 1	CTT A Leu S 845	GT # Ser 1	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	SEL	AAA 1 Lys 1 855	TAC (Tyr \	GTA (/al /	GAT / Asp /	Asn (CAA A Gln A 860	AGA T Arg L	TA I	TA :	TCT Ser	2592
ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	TAT Tyr	ATT . Ile : 870	AAG Lys										2613

(2) INFORMATION FOR SEQ ID NO: 8:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- 52 -

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

tor Dro Dhe Val Asp Ive Gla Dhe Asp Tur Ive Asp Dec Val Asp

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 . 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 455 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495

Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 510

Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu
530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 580 585

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

- 54 -Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys (2) INFORMATION FOR SEQ ID NO: 9: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 2628 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear (ii) MOLECULE TYPE: DNA (genomic) (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION:1..2628 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9: ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

20

GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

45

GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG GAA

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

50

65	5		Va	I FI	70)	c ly	r ly	I AS	T TC	r Th	т Ту	r Le	eu S	er	Thr 80	240
Yal	, na		u Dy	89	5	· Iyı	. Let	u Ly	9(•	. Th	r Ly	s Le	u P	he 95	Glu	288
ALG	1 116	e ly.	100)	. vat	. Leu	i Gi)	10	g met 5	G CTC	Le	u Th	r Se 11	r II 0	le '	Val	336
Arg	, (1)	119	5	FILE	LIP	GIY	120)	r Thi	: ATT	Asp	12:	r Gl	u Le	eu I	Lys	384
Vai	130)	, 1111	. nsi	cys	135	ASI	ı vaı	Lie	CAA Gln	Pro 140	Ası	Gl;	y Se	rj	yr	432
145	261	GIU	GIU	Leu	150	ren	Val	116	: Ile	GGG Gly 155	Pro	Ser	Ala	a As	р I 1	le .60	480
116	GIII	Pile	GIU	165	Lys	ser	Pne	Gly	170		Val	Leu	Asr	1 Le	u T 5	hr	528
ALG	ns:	Gly	180	GIY	SEL	Inr	GIN	197 185	11e	CGT Arg	Phe	Ser	Pro 190	As	p P	he	576
1111	FIIE	195	PHE	GIU	GIU	ser	200	GIu	Val	GAT Asp	Thr	Asn 205	Pro	Let	ı L	eu	624
Gly	210	GLY	пуз	Pne	ATA	215	Asp	Pro	Ala	GTG Val	Thr 220	Leu	Ala	His	3 G.	lu	672
225	116	urs	ATA	GIĀ	230	Arg	ren	Tyr	Gly	ATT Ile 235	Ala	Ile	Asn	Pro	24	sn 10	720
Arg	Val	Pne	гуя	245	ABI	Inr .	asn	Ala	Tyr 250		Glu	Met	Ser	Gly 255	Le	eu	768
Giu	Val	SET	260	GIU	GIU .	Leu .	Arg	7hr 265	Phe	GGT (Gly	His	Asp 270	Ala	Ly	'6	816
TTT . Phe	TTE	GAC Asp 275	AGC Ser	TTG Leu	CAG (Gln (stu 4	AAC Asn 280	GAG Glu	TTC Phe	CGT (Arg)	Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AA As	C n	864
	290	гÀв	Asp	TT6	ALA S	ser 1 295	rnr :	Leu .	Asn :	Lys ;	Ala 300	Lys	Ser	Ile	Va	1	912
GGT A Gly 7 305	ACC . Thr	ACT Thr	GCT (Ala :	ser !	TTA (Leu (310	CAG 1	rat : Cyr I	ATG . Met :	Lys i	AAT (Asn V 315	TT '	TTT Phe	AAA Lys	GAG Glu	AA Ly 32	S	960
TAT (CTC (Leu)	CTA :	ser (GAA (Glu / 325	Asp I	CA Thr S	CT (Ser (Gly :	AAA : Lys 1 330	ITT 1 Phe S	CG (Ser)	GTA Val	Asp	AAA Lys 335	TT.	A u	1008

-			34	0	y.	- 2,3	net.	349	5	r G11	u Il	е Ту	r Th	r Gl 0	G GAT u Asp	
		35	5		- 1110	- Dys	360	. rec	I ASI	1 Arc	g Ly	s Th	r Ту 5	r Le	G AAT u Asn	1104
	370)			116	375	ile	ASI	1116	val	380	o Ly	s Va	l As:	T TAC	1152
385		,.	. 7.0]	J GI	390	, ven	Leu	Arg	ASN	395	Ası	ı Lei	ı Ala	a Ala	A AAC A Asn 400	1200
		. 01,	, U11	405		GIU	ııe	ASI	410	Met	Asr	1 Phe	? Thi	415		1248
-,-			420)	Deu	FIIC	GIU	425	Tyr	Lys	Leu	Leu	430	Val	A AGA Arg	1296
- -,		435		261	БУБ	1111	440	ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	AAG Lys	1344
-	450	nop	Gly	ALG	Den	455	АБР	reu	Cys	Ile	Lys 460	Val	Asn	Asn	TGG	1392
465	500	1110	1116	361	470	TCA Ser	GIU	Asp	Asn	Phe 475	Thr	Asn	Asp	Leu	As n 480	1440
-,-	O1,	0.10	GIU	485	****	TCT Ser	Asp	Inr	490	He	Glu	Ala	Ala	Glu 495	Glu	1488
	***	Jer	500	vaħ	Deu	ATA Ile	GTÜ	505	Tyr	Tyr	Leu	Thr	Phe 510	Asn	Phe	1536
p		515	110	Giu	VBII		520	TIG	GIu	Asn	Leu	Ser 525	Ser	Asp	Ile	1584
	530	01 11	neu	GIU	beu	ATG Met 535	PIO.	ABN	TTE	Glu	Arg 540	Phe	Pro	Asn	Gly	1632
545	ay e	TYL	GIU	nea	550	AAA ' Lys '	ıyr :	rnr	Met	Phe 555	His	Tyr	Leu	Arg	Ala 560	1680
	-		GIU	565	Gry	AAA ' Lys :	ser /	arg	570	Ala .	Leu	Thr	Asn	Ser 575	Val	1728
Aou	914	vra	580	nen	АВП .	CCT :	ser /	85 585	Val '	Tyr '	Thr	Phe	Phe 590	Ser	Ser	1776
GAC Asp	TAT	GTA Val 595	AAG Lys	AAA Lys	GTT : Val :	AAT 1 Asn 1	AAA (Lys /	SCT A	ACG (Thr (GAG (Glu /	Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

WO 98/07864 PCT/GB97/02273

GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA ACT AGC GAA Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 620	1872
GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT ATT CCA TAT Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 630 635 640	1920
ATA GGA CCT GCT TTA AAT ATA GGT AAT ATG TTA TAT AAA GAT GAT	1968
GTA GGT GCT TTA ATA TTT TCA GGA GCT GTT ATT CTG TTA GAA TTT ATA Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile 660 665 670	2016
CCA GAG ATT GCA ATA CCT GTA TTA GGT ACT TTT GCA CTT GTA TCA TAT Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr 675 680 685	2064
ATT GCG AAT AAG GTT CTA ACC GTT CAA ACA ATA GAT AAT GCT TTA AGT Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 695 700	211?
AAA AGA AAT GAA AAA TGG GAT GAG GTC TAT AAA TAT ATA GTA ACA AAT Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn 710 715 720	2160
TGG TTA GCA AAG GTT AAT ACA CAG ATT GAT CTA ATA AGA AAA AAA ATG Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Met 725 730 735	2208
AAA GAA GCT TTA GAA AAT CAA GCA GAA GCA ACA AAG GCT ATA ATA AAC Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn 740 745	2256
TAT CAG TAT AAT CAA TAT ACT GAG GAA GAG AAA AAT AAT ATT AAT TTT Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe 755 760 765	2304
AAT ATT GAT GAT TTA AGT TCG AAA CTT AAT GAG TCT ATA AAT AAA GCT Asn lle Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 770 780	2352
ATG ATT AAT AAA AAA TTT TTG AAT CAA TGC TCT GTT TCA TAT TTA Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu 785 790 795	2400
ATG AAT TCT ATG ATC CCT TAT GGT GTT AAA CGG TTA GAA GAT TTT GAT Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp 805 810 815	2448
GCT AGT CTT AAA GAT GCA TTA TTA AAG TAT ATA TAT GAT AAT AGA GGA Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820 825	2496
ACT TTA ATT GGT CAA GTA GAT AGA TTA AAA GAT AAA GTT AAT ACA Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 835 840 845	2544
CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA GAT AAT CAA Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln 850 855	2592
AGA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys + 865 870 875	2628

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 876 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys
260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp
450 455 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile
515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 530 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 555 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser 585

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu
610 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 655

WO 98/07864 PCT/GB97/02273

- 60 -

Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr 680 Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser 690 Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala 775 Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu

Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp

Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820

Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr

Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2637 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE_TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
- ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT 48 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
- GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG 96 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 25

		35		AAG Lys		r	40	iys I	ie T	rp v	al I	le P 45	ro G	lu	Arg	144
1.20	50			AAC Asn		55	ilu G	iy A	sp L	eu A	sn Pi 60	ro P	ro P	ro	Glu	192
65	2,5			CCA (70	er 1	yr 1	yr A	sp Se	er 11 75	nr Ty	r Le	eu S	er	Thr 80	240
	******	014	5 ,5	GAT ASP A		yr D	eu L	ys G	y va 90	al Tr	ır Ly	s Le	u Pl	ne (Glu	288
5		-,-	100	ACT C	JP I	cu G.	10	19 ME	t Le	u Le	u Th	r Se	r Il 0	.e '	Val	336
5	,	115		TTT T Phe T	rp G	12	20 20	r In	r ii	e As	p Th	r Gl	u Le	u I	Lys	384
	130			AAC T Asn C	1.	35	ın va	1 11	e GI:	n Pr	o Ası O	o Gl	y Se	r 1	γr	432
145					50	su va		e 11	15.	y Pro	Sez	: Ala	a As	P I 1	le 60	480
			1	GC AL Cys Ly .65	, 9 36	I PN	e GI	170	GII	ı Val	Leu	Asr.	1 Let	ıT.	hr	528
		1	180	GC TO	- 11.	ir Gi	189	2 116	e Arg	Phe	Ser	Pro 190	Asp	Pi	he	576
	1	95		AG GA lu Gl	u 36	200)	ı vaı	Asp	Thr	Asn 205	Pro	Leu	Le	u	624
	10	-, -	.,	ii al	21	5	PIC	Ala	Val	7hr 220	Leu	Ala	His	Gl	.u	672
CTG A Leu I 225			ZG G.	23		a nen	ııyr	GIY	235	Ala	Ile	Asn	Pro	Ав 24	n 0	720
CGC G Arg V			24	15	1 111	ASD	Ala	250	Tyr	Glu	Met	Ser	Gly 255	Le	u	768
GAA G Glu V		2	60	.u G1(. Dec	. Arg	265	Pne	GIA	Gly	His	Asp 270	Ala	Ly	8	816
TTT AT	2	75	ar ne	u 911	GIC	280	GIU	Pne	Arg	Leu	Tyr 285	Tyr	Tyr	Ası	n	864
AAG TY Lys Pl 29	TT AU ne L _l 90	AA GA /S As	AT AT Sp Il	T GCA e Ala	AGI Ser 295	inr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GT(Val	3 L	912

309	5		- 710	. Je.	310)	ı ıyı	r mei	: Lys	315	ı Val	l Pho	e Ly	s Gl	G AAA u Lys 320	960
TYI	. ne	ı re	ı ser	325	Ast	Inr	Ser	: Gly	/ Lys	Phe	Ser	'Va	l Ası	33!		1008
AA# Lys	TTI Phe	GA:	Lys 340	, ner	TAC Tyr	AAA Lys	ATC Met	Leu 349	Thr	GAG Glu	ATT	TAC Tyl	C ACA Thi 350	: Glu	GAT Asp	1056
AAT Asn	TTI	GT7 Val 355	r rys	TTI Phe	TTI Phe	AAA Lys	GTA Val 360	Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	Tyr	TTC Lev	AAT Asn	1104
FIIC	370	, nys	, Ala	. vai	Pne	375	TTE	Asn	Ile	Val	Pro 380	Lys	Val	Asn	TAC	1152
385		Lyz	GAT Asp	Gly	390	ASN	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400	1200
	Aan	Gly	CAA Gln	405	Inr	GIU	TTE	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
Dya	Vali	FIIC	ACT Thr 420	GIY	Leu	Pne	GIU	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
Gly	116	435	ACT	ser	гув	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
116	450	GIY	CGT Arg	Cys	Asp	455	Ala	Leu	Asn	Asp	Leu 460	Cys	Ile	Lys	Val	1392
465	VSII	110	GAC Asp	Leu	470	Pne	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	Asn 480	1440
vah	Dea	ABII	AAA Lys	485	GIU	GIU	11e	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495	Ala	1488
Ala	GIU	GIU	AAT Asn 500	ııe	ser	ren	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	Thr	1536
Pne	ASN	515	GAT Asp	ABN	GIU	Pro	Glu 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Leu	Ser	1584
261	530	116	ATA Ile	GIÀ	GIN	535	GIU	Leu	Met	Pro	Asn 540	Ile	Glu	Arg	Phe	1632
545	MSII	GIY	AAA Lys	гÀа	550	GIU	Leu	Asp	Lys	Tyr 555	Thr	Met	Phe	His	Tyr 560	1680
CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT . Ser .	AGG . Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728

									•				
			580				585	FIO S	er Arg		Tyr Ti 590	ir Phe	1776
TT'	TC1	TCA Ser 595	GAC Asp	TAT (TA AAG	G AAA B Lys 600	GTT I	AAT AA Asn L	AA GC1 ys Ala	ACG (Thr (GAG GC Glu Al	A GCT a Ala	1824
ATC Met	Phe 610	TTA Leu	GGC Gly	TGG G Trp V	TA GAA al Glu 615	CAA Gln	TTA (STA TA /al Ty	AT GAT 'r Asp 620	TTT A	ACC GA	T GAA p Glu	1872
625				6	30	nop	mys I	63	a Asp 5	ATA A Ile T	hr Ile	e Ile 640	1920
ATT Ile	CCA Pro	TAT Tyr		GGA CO Gly P: 545	CT GCT ro Ala	TTA . Leu .	wen T	TA GG le Gl 50	T AAT y Asn	ATG T	TA TAT eu Ty: 655	: Lys	1968
GAT Asp	GAT Asp	TTT Phe	GTA G Val G 660	GT GG	TTA La Leu		Phe S	CA GG er Gl	A GCT y Ala	GTT AT Val II	TT CTC le Leu 70	TTA Leu	2016
		675				680	LO V	ar re/	1 GIA	ACT TT Thr Ph	e Ala	Leu	2064
GTA Val	TCA Ser 690	TAT I	ATT G Ile A	CG AA la As	T AAG n Lys 695	GTT C	TA AC	CC GT1	CAA Gln 700	ACA AT Thr Il	A GAT e Asp	AAT Asn	2112
705			-,	71	0	nys 1	Th We	715	Val '	TAT AA Tyr Ly	s Tyr	Ile 720	2160
			72	25	,0	VOL A	73	r Gin O	Ile A	GAT CTA	u Ile 735	Arg	2208
	_	7.	40			74	15	n Ala	Glu A	GCA ACA	Lys	Ala	2256
	7	55	•	,-	,	760	/4 IIII	GIU	Glu G	AG AAA lu Lys 65	Asn .	Asn	2304
7	70				775	cu se	ı sei	Lys	Teu A	AT GAG sn Glu	Ser :	Ile	2352
785	•			790	,	ын шу	s rue	795	Asn G	AA TGC ln Cys	Ser 7	/al 300	2400
			805	5	1	TE PI	810	GIÀ	Val L	AA CGG ys Arg	Leu 0	Slu	2448
GAT T	IT GI he As	AT GC sp Al 82		CTT Leu	AAA G Lys A	AT GCI Sp Ala 825	a Leu	TTA L	AAG TA Lys T ₎	AT ATA	TAT C	AT Lsp	2496
AAT AG Asn A	GA GC Fg G1 83		T TTA r Leu	ATT	U-7 G.	AA GTA ln Val	A GAT L Asp	AGA :	TTA AF Leu Ly 84	AA GAT 's Asp 5	AAA G Lys V	TT al	2544

WO 98/07864 PCT/GB97/02273

- 64 -

AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TAA *		2637
(2)	INFO	RMAT	CION	FOR	SEQ	ID 1	10: 1	.2:								
					~		·									

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 879 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

- Glu Val. Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435
 440
 445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450
- Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480
- Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495
- Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505
- Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525
- Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 535 540
- Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 555 560
- Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575
- Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590
- Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605

- Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620
- Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640
- Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875
- (2) INFORMATION FOR SEQ ID NO: 13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2862 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2862
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATG CAG TTC GTG Met Gln Phe Val 1	AAC AAG CAG TTC Asn Lys Gln Phe 5	AAC TAT AAG GAC CCT GTA AAC GGT Asn Tyr Lys Asp Pro Val Asn Gly 10	48
GTT GAC ATT GCC Val Asp Ile Ala 20	TAC ATC AAA ATT Tyr Ile Lys Ile	CCA AAC GCC GGC CAG ATG CAG CCG Pro Asn Ala Gly Gln Met Gln Pro 25	96
GTG AAG GCT TTC Val Lys Ala Phe 35	AAG ATT CAT AAC Lys Ile His Asn 40	AAA ATC TGG GTT ATT CCG GAA CGC Lys Ile Trp Val Ile Pro Glu Arg 45	144
50	55	GGA GAC TTG AAC CCG CCG CCG GAA Gly Asp Leu Asn Pro Pro Pro Glu 60	192
65	70	TAC GAT TCA ACC TAT CTG AGC ACA Tyr Asp Ser Thr Tyr Leu Ser Thr 75 80	240
	85 THE TYPE DEC 1	AAG GGA GTG ACC AAA TTA TTC GAG Lys Gly Val Thr Lys Leu Phe Glu 90 95	288
100	in the field Gly	CGT ATG CTG CTG ACC TCA ATC GTC Arg Met Leu Leu Thr Ser Ile Val 105	336
115	120	AGT ACC ATT GAC ACG GAG TTG AAG Ser Thr Ile Asp Thr Glu Leu Lys 125	384
130	135	GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Gln Pro Asp Gly Ser Tyr 140	432
145	150	TC ATC GGG CCC TCC GCG GAC ATT le lle Gly Pro Ser Ala Asp Ile 155	480
16	5	GC CAC GAA GTG TTG AAC CTG ACG ly His Glu Val Leu Asn Leu Thr 170	528
180	11	AC ATT CGT TTC AGC CCA GAC TTC yr Ile Arg Phe Ser Pro Asp Phe 190	576
195	200	AG GTT GAT ACC AAC CCG CTG TTG Lu Val Asp Thr Asn Pro Leu Leu 205	624
210	215	CA GCG GTG ACC CTG GCA CAC GAG TO Ala Val Thr Leu Ala His Glu 220	672
225	230 Alg Let 1y	T GGC ATT GCG ATT AAC CCG AAC T Gly Ile Ala Ile Asn Pro Asn 235	720
245	i min min min min	C TAC TAC GAG ATG AGT GGT TTA a Tyr Tyr Glu Met Ser Gly Leu 250	768
GAA GTA AGC TTC GAG Glu Val Ser Phe Glu 260	GAA CTG CGC ACC Glu Leu Arg Th 26	G TTC GGT GGC CAT GAT GCG AAG r Phe Gly Gly His Asp Ala Lys 270	816

TTT Phe	ATC Ile	GAC Asp 275	ser	TTC Leu	CAG Gln	GAG Glu	AAC Asn 280	Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	Tyr	TAC	AAC Asn	864
AAG Lys	TTT Phe 290	гÀв	GAT Asp	ATI	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	Lys	TCC	ATT	GTG Val	912
GGT Gly 305	inr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	GIn	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT	GAT Asp	AAG Lys 340	TTA Leu	TAC	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	TAA Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
Pne	Asn	GIÀ	GIn	405	Thr	GAA Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	1248
Lys	Asn	Phe	Thr 420	Gly	Leu	TTT Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
ATC Ile	GAA Glu 450	GGT Gly	CGT Arg	TGC Cys	GAT Asp	GGG Gly 455	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 460	TGT Cys	ATC Ile	AAA Lys	GTT Val	1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 485	GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 490	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 495	GCA Ala	1488
GCA Ala	GAA Glu	GAA Glu	AAT Asn 500	ATT Ile	AGT Ser	TTA Leu	Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	1536
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA Glu	CCT Pro	GAA Glu 520	TAA Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

5	45		•	•	5	50	u .	icu i	rap	гÀв	555	In	T AT	t Ph	ne H	lis	Tyr 560	1680
		•		5	65		, Lu .,	.15	, r y	570	ser	Arg	G AT	e Al	a L 5	eu 75	Thr	1728
			5	80			icu D	5	85	PIO	ser	Arc	r GTT Val	. Ty 59	r T	hr	Phe	1776
		5	95		2		6	00	aı,	4SN	гÀа	Ala	ACG Thr 605	Gl	u Al	la	Ala	1824
	61	.0		, -	- -	6	15	r11 Te	eu v	aı	lyr	Asp 620		Thi	r As	P	Glu	1872
62	5				63	0	it As	չը եչ	/S I	ie.	635	Asp	ATA Ile	Thi	: Il	e :	Ile 540	1920
		- 2		64	5	·	a De	u As	6	50	этÀ	Asn	ATG Met	Leu	65	r 1 5	Lys	1968
•	•		66	0	,		·	66	5	er (эту ,	Ala	GTT Val	11e 670	Le	n I	eu	2016
		67	5			- 11	68	0	o va	at f	eu (Sly	ACT Thr 685	Phe	Ala	a L	eu	2064
	690) -				69	5	r ne	u II	IF V	al	700	ACA Thr	Ile	Asp) A:	sn	2112
705			3		710		ı Dys	, 111	AS	р G 7	10 V	al	TAT . Tyr :	Lys	Тух	7:	le 20	2160
				725	5	y-	, vai	wat	73	T G.	ın I	le i	GAT (Asp 1	Leu	Ile 735	Aı	g	2208
•	•		740			200	GIU	745	GI	n A.	la G	lu)		Thr 750	Lys	Al	.a	2256
		755	-7-		-/-	7.51	760	ıyı	I'M;	r Gi	lu G	lu c	GAG A Glu L 765	ys .	Asn	As	n	2304
	770				щ	775	neu	Ser	261	: гу	's Le 78	eu A 30	AT G Asn G	lu s	Ser	Il	е	2352
785	•				790		ven	nys	Pne	79	u As 5	ın G	AA T	ys S	Ser	Va:	1 0	2400
TCA '	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GI	T GI y Va	T A	AA C	rg L	TTA Leu 115	GA/ Glu	1 V	2448

GAT Asp	TTT	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	2688
TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	2736
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	2784
TGC Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	GAG Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	2832
CTC Leu 945	AAG Lys	CCT Pro	GCC Ala	AAG Lys	TCA Ser 950	GCT Ala	GAA Glu	GCT Ala	TAG *							2862

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

- Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
- Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
- Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
- Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
- Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
- Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
- Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220
- Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 235 230
- Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
- Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 350
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435
 440
 445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 460

						•	, 0					4 75	•				r Asn 480
											490	,				49	
				-						503)				510)	1 Thr
				-					520					529	5		ı Ser
							•	,,,					540	l			, Phe
						٠.						222					Tyr 560
	Leu	Arg	Ala	a Gl	n Gl 56	u Pł 5	ie G	lu	His	Gly	Lys 570	Ser	Arg	Ile	Ala	Leu 575	Thr
				-	•					202					590		Phe
									800					605			Ala
					•		•	10					620				Glu
						0,5	U					Ala 635					640
					04.	•					650	Gly				655	
				000	,					005		Gly			670		
								,	000	•		Leu		685			
		0,50					03	70				Val	700				
						, _ \	•					Glu 715					720
					/23						730	Gln				735	
				/40					Ī	/45		Ala			750		
I	le :	Ile	Asn 755	Tyr	Gln	Туз	: Ae	n G	in 1	[yr	Thr	Glu	Glu	Glu	Lys	Asn	Asn

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu

805

WO 98/07864 PCT/GB97/02273

- 73 -

	p Phe	Asp	Ala 820	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	Ile 830		Asp	
Ası	n Arg	Gly 835	Thr	Leu :	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys. 845			Val	
			Leu		`						860					
			Arg							8/5					880	
									990					895		
			Gly 1 900					3 03.					910			
			Ser A			-	20					925				
Cys	Phe 930	Arg	Ser (ys A	sp L 9	eu A 35	arg i	Arg	Leu	Glu	Met 940	Tyr	Cys	Ala	Pro	
Leu 945	Lys	Pro .	Ala L		er A 50	la G	lu 1	Ala	*							
(2)	INFO	RMAT	ION F	OR SI	EQ II	ON C	: 15	5 :								
	(ii) (ix)	(E) (D) MOLE	LENG TYP STR TOPG	E: nu ANDEI OLOGY	nclei NESS : li	ic a S: do .nea:	cid oubl r	.e				٠				
		(A) (B)	NAME LOCA	TION	:1	2724										
	(xi)	(A) (B) SEQU	NAMI LOCA ENCE	TION DESC	:1 RIPT	2724 ION:	SE									
ATG ((xi) CAG T Gln E	(A) (B) SEQU TC G	NAME LOCA ENCE TG AA al As	DESC C AA n Ly:	RIPT G CA	2724 ION: G TI n Ph	SEC A	AC T.	AT A yr L 10	AG G ys A	sp P	ro V	al A	sn G 15	ly	48
ATG (Met ((xi) CAG 1 Gln E GAC A	(A) (B) SEQU TC G Phe V	NAME LOCA ENCE TG AA al As CC TA la Ty	DESC C AA n Ly: 5	RIPT G CA S Gl:	ION: G TT n Ph A AT	SEC AL	AC T. sn T	AT A yr L 10 AC G	AG G ys A CC G la G	sp P GC C ly G	ro V AG A' ln Me	al A TG C et G	sn G 15 AG C ln P	CG ro	48 96
ATG (Met (Control of the Control of	(xi) CAG T Gln E GAC A Asp I AG G	(A) (B) SEQU TTC G TT G (le A) CT TT la Pi	NAME LOCAL ENCE TG AM al As CC TA la Ty 20 CC AM CC Ly:	DESC C AAG n Ly: S C ATC r Ile	RIPT G CA S Gl: C AAA E Ly: C CA His	ION: G TT n Ph A AT s Il	C AM	AC T. Sn T. CA A. CO A: 25	AT A yr L 10 AC G sn A	AG G. CC GG la G. GG GT TP Va	SC Cally G.	AG A' ln Me : TT CC le Pi	rG C et G 30	sn G 15 AG C ln P	CG ro GC rg	
ATG (Met (Control of the Met (Control of the M	(xi) CAG TASP I AG G YS A CA T	(A) (B) SEQU TC G The V TT G le A CT TT la Pi 35	NAME LOCAL TO A A A A A A A A A A A A A A A A A A	DESC: C AAA n Ly: 5 C ATC r Ile G ATT s Ile	RIPT G CAG S Gl: C AA; C His GGA GGI GS GGA GS	ION: G TT n Ph A AT S II	T CO e Pr C AA n Ly 0	AC TI	AT A yr L 10 AC G sn A CC TC Le T AC TI AC AC TI AC	AG G. YS A. CC GG la G: TG AF	SP P GC Ci ly G. TT A: AC CC sn P:	AG AMIN MONTH COLOR PROPERTY ASSESSMENT OF THE PROPERTY ASSESSMENT	rg C. et G. 30 cg G. co G.	Sn G 15 AG C ln P AA C lu A	CG ro GC rg	96
ATG (Met (Control of the Control of	(xi) CAG T GIN E ASP I ASP A ASP A CA T T T T T T T T T T T T T T T T T T T	(A) (B) SEQU TC G Phe V TT G le A CT T: la Pi 35 TT AC he Tr	NAME LOCAL TO A A A A A A A A A A A A A A A A A A	DESC: C AAA n Ly: 5 C ATC r Ile G ATT s Ile C CCG n Pro A GTT 70	RIPT G CA S Gl: C AA His GGA GGU SS TCA Ser	ION: G TT n Ph A AT S II C AA I GA I GI I TAC I TY	SEC AMERICAN A GG	AC TAN TO AN TAN TAN TAN TAN TAN TAN TAN TAN TAN	AT A AYY L AC GARA AC TO AC	AG G. YS A. CC GG la G: GG AA CC Th	SC CALL COST PROSE	AG AMILE COLOR PROPERTY COLOR PROPER	rg C. et G 30 cg G co G co P cg CC	sn G 15 AG C ln P AA C lu A	CG CG CG CG CG CA Lu CA CA	96 144

CG Ar	T AT g Il	T TA e Ty	T TC: r Se: 10		T GAG	CTC Lev	GG(C CG: / Arc	g Mei	G CTO	G CTO	G ACC	C TC: Se:	r Il	C GTC e Val	:	336
CG(Ar	GG Gl	A ATO		A TTT	TGC Tr	GGT Gly	GGC Gly 120	Sei	T ACC	T Ile	GA(C ACC Thi	: Glu	G TT	G AAG u Lys		384
	130)			. Cys	135	ASI	val	. 116	GIR	140) Asp	Gly	/ Se	TAC Tyr		432
145		. 010	. 010	. Deu	150	. Leu	val	TTE	: Ile	155	Pro	Ser	Ala	Asp	ATT Ile 160		480
		• • • • • • • • • • • • • • • • • • • •	. 010	165	Lys	Ser	Pne	GIÀ	170	Glu	Val	Leu	Asn	Leu 175			528
3	7.01.	. Gly	180	GIY	261	Ini	GIR	185	ııe	Arg	Phe	Ser	Pro 190	Asp	TTC Phe		576
		195	File	Giu	GIU	Ser	200	GIU	vai	Asp	Thr	Asn 205	Pro	Leu		•	624
Gly	210	GIY	пуз	Pne	AIA	215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His		l	672
225	116	*****	nia	GIY	230	Arg	Leu	Tyr	GIY	11e 235	Ala	ATT Ile	Asn	Pro	Asn 240	7	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	7	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	8	16
FILE	116	275	ser	Leu	GIN	GIU	Asn 280	Glu	Phe	Arg	Leu	TAC Tyr 285	Tyr	Tyr	Asn	8	64
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	9	12
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	9	60
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT : Ser	Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	10	08
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA / Lys	Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	10	56
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT . Phe	ras ,	GTA (Val : 360	CTT . Leu .	AAC . Asn .	AGA . Arg	Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	11	04

TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA AAT Phe Asp Lys Ala Val Ph Lys Ile Asn Ile Val Pro Lys Val Asn 370 375 380	TAC 1152 Tyr
	Asn 400
TTT AAT GGT CAA AAT ACA GAA ATT AAT AAT ATG AAT TTT ACT AAA (Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys I 405 410 415	Leu
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT GTA F Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val F 420 425 430	Arg
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AAT AGIY Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn L	ys
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AAA G Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys V 450 455 460	al
	sn 80
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GAA GG Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ai 485 490 495	la
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA ACAA Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Th	ır
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT TC Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Se 515 520 525	er
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA TT Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Ph 530 535 540	e
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT TAPPRO Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Ty: 545 550 555	r 0
CTT CGT GCT CAA GAA TTT GAA CAT GGT AAA TCT AGG ATT GCT TTA ACL Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thi 565 570 575	r
AAT TCT GTT AAC GAA GCA TTA TTA AAT CCT AGT CGT GTT TAT ACA TTT Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590	•
TTT TCT TCA GAC TAT GTA AAG AAA GTT AAT AAA GCT ACG GAG GCA GCT Phe Ser Ser Asp Tyr Val Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605	•
ATG TTT TTA GGC TGG GTA GAA CAA TTA GTA TAT GAT TTT ACC GAT GAA Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620	ı
ACT AGC GAA GTA AGT ACT ACG GAT AAA ATT GCG GAT ATA ACT ATA ATT Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 640	1920

ATT Ile	CCA Pro	TAT Tyr	TATA	GGA Gly 645	Pro	GCT Ala	TTA Leu	AA1 Asn	ATA Ile 650	Gly	AAT Asn	ATC Met	TTA Lev	TAT Tyt	AAA Lys	1968
GAT Asp	GAT Asp	TTI Ph	GTA Val 660	GIA	GCT Ala	TTA Leu	ATA Ile	Phe	Ser	GGA Gly	GCT Ala	GTI Val	Ile 670	Leu	TTA Leu	2016
GAA Glu	TTT Phe	ATA Ile 675	Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT	2064
GTA Val	TCA Ser 690	Tyr	ATT	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	Thr	ATA Ile	GAT Asp	AAT Asn	2112
705	reu	ser	AAA Lys	Arg	710	Glu	Lys	Trp	Asp	Glu 715	Val	Tyr	Lys	Tyr	Ile 720	2160
vai	Thr	Asn	TGG Trp	Leu 725	Ala	Lys	Val	Asn	Thr 730	Gln	Ile	Asp	Leu	Ile 735	Arg	2208
rya	гÀа	Mec	AAA Lys 740	GIU	ATA	Leu	Glu	745	Gln	Ala	Glu	Ala	Thr 750	Lys	Ala	2256
ile	116	755	TAT Tyr	Gin	Tyr	Asn	760	Tyr	Thr	Glu	Glu	Glu 765	Lys	Asn	Asn	2304
TIE	770	Pne	AAT Asn	116	Asp	775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile	2352
785	riys	Ala	ATG Met	Ile	790	Ile	Asn	Lys	Phe	Leu 795	Asn	Gln	Сув	Ser	Val 800	2400
Ser	Tyr	Leu	ATG Met	Asn 805	Ser	Met	Ile	Pro	Tyr 810	Gly	Val	Lys	Arg	Leu 815	Glu	2448
Asp	Phe	Asp	GCT Ala 820	Ser	Leu	Lys	Asp	Ala 825	Leu	Leu	Lys	Tyr	Ile 830	Tyr	Asp	2496
Asn	Arg	61y 835	ACT Thr	Leu	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	Asp	Lys	Val	2544
Asn	Asn 850	Thr	CTT Leu	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	Tyr	Val	2592
GAT Asp 865	TAA Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	2640
CCT Pro	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA Gln	ATA Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	CAA Gln	TCT Ser 895	GAT Asp	2688
ATT Ile	GAT Asp	ACA Thr	CAT His 900	AAT Asn	AGA Arg	ATT Ile	Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA *				,	2724

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 908 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr
130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys 650

WO 98/07864 PCT/GB97/02273

- 79 -

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 740 745 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 865 870 875 880

Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 885 890 895

Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu * 900 905

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3042
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1

48

GTT Val	GAC Asp	ATT Ile	GCC Ala 20	Tyr	ATC Ile	Lys	ATT Ile	CCA Pro 25	Asn	GCC Ala	GGC	CAG Gln	ATG Met 30	Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Fue	Lys	ATT Ile	CAT His	AAC Asn 40	Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC	144
GAT Asp	ACA Thr 50	hve	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
65	гÀа	GIN	vaı	Pro	70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	80	240
Asp	Asn	GIU	гÀЗ	Asp 85	AAC Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu	288
Arg	11e	Tyr	Ser 100	Thr	GAC Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864

AAG TTT AAA GAT ATT GCA AGT ACA CTG AAC AAG GCT AAG TCC Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 290 295 300	Ile Val	912
GGT ACC ACT GCT TCA TTA CAG TAT ATG AAA AAT GTT TTT AAA Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys 305 310 315	Glu Lys 320	960
TAT CTC CTA TCT GAA GAT ACA TCT GGA AAA TTT TCG GTA GAT Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp 325	Lys Leu 335	1008
AAA TTT GAT AAG TTA TAC AAA ATG TTA ACA GAG ATT TAC ACA Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 340 345	Glu Asp	1056
AAT TTT GTT AAG TTT TTT AAA GTA CTT AAC AGA AAA ACA TAT : Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr 1 355 360 365	Leu Asn	1104
TTT GAT AAA GCC GTA TTT AAG ATA AAT ATA GTA CCT AAG GTA APPHE ASP Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val APPHE ASP 370	Asn Tyr	1152
ACA ATA TAT GAT GGA TTT AAT TTA AGA AAT ACA AAT TTA GCA G Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala A 395	la Asn 400	1200
	ys Leu 15	1248
AAA AAT TTT ACT GGA TTG TTT GAA TTT TAT AAG TTG CTA TGT G Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys V 420 425	al Arg	1296
GGG ATA ATA ACT TCT AAA ACT AAA TCA TTA GAT AAA GGA TAC AA Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr As 435 440 445	sn Lys	1344
ATC GAA GGT CGT TGC GAT GGG GCA TTA AAT GAT TTA TGT ATC AA Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Ly 450 460	's Val	1392
AAT AAT TGG GAC TTG TTT TTT AGT CCT TCA GAA GAT AAT TTT AC Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Th 470 475	r Asn 480	1440
GAT CTA AAT AAA GGA GAA GAA ATT ACA TCT GAT ACT AAT ATA GA Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Gl 485 490	u Ala 5	1488
GCA GAA GAA AAT ATT AGT TTA GAT TTA ATA CAA CAA TAT TAT TTA Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu 500 510	1 Thr	1536
TTT AAT TTT GAT AAT GAA CCT GAA AAT ATT TCA ATA GAA AAT CTT Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu 515 520 525	Ser	1584
AGT GAC ATT ATA GGC CAA TTA GAA CTT ATG CCT AAT ATA GAA AGA Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg 530 535	Phe	1632
CCT AAT GGA AAA AAG TAT GAG TTA GAT AAA TAT ACT ATG TTC CAT Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His 545 550 555	TAT Tyr 560	1680

CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT	GGT Gly	AAA Lys 570	Ser	AGG Arg	ATT	GCT Ala	TTA Leu 575	ACA Thr	1728
AAT Asn	TCT	GTT Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	Pro	AGT Ser	CGT	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776
TTT Phe	TCT Ser	TCA Ser 595	GAC Asp	TAT Tyr	GTA Val	AAG Lys	AAA Lys 600	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala	1824
Met	Phe 610	Leu	Gly	Trp	Val	Glu 615	Gln	Leu	Val	Tyr	Asp 620	Phe	Thr	Asp		1872
Thr 625	Ser	Glu	Val	Ser	Thr 630	Thr	Asp	Lys	Ile	Ala 635	GAT Asp	Ile	Thr	Ile	Ile 640	1920
ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT Gly	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu	2016
GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	2064
GTA Val	TCA Ser 690	TAT Tyr	ATT Ile	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA Ile	GAT Asp	AAT Asn	2112
GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	ÀAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	TAA Asn	AAT Asn	2304
ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GAT Asp	2496

		83	5			- 01)	840)	. Asp	Arg	, ren	845	Asp	Ly:	A GTT s Val	2544
AA As	T AA' n As: 85		A CT r Le	T AG: u Sei	r ACA	GAT Asp 855	, 116	CCI Pro	TTI Phe	CAG Gln	CTT Leu 860	Ser	Lys	TAC	GTA Val	2592
GA' Asj 86	T AA' p Ası 5	r CA	A AGA	A TTA J Leu	TTA Leu 870	SET	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	Tyr	ATT Ile	AAG Lys	TCA Ser	GGC Gly 880	2640
				885	770	nia	uis	ıyr	890	Gin	His	Asp	Glu	Ala 895		2688
		,.	900		Lys	GIU	GIN	905	Asn	Ala	Phe	Tyr	Glu 910	Ile	Leu	2736
		915		TTA Leu	Adii	GIU	920	GIN	Arg	Asn	Ala	Phe 925	Ile	Gln	Ser	2784
TTA Leu	Lys 930		GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys	2832
AAG Lys 945		AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960	2880
			7.511	GCG Ala 965	FIIC	TYL	GIU	TTE	970	His	Leu	Pro .	Asn	Leu 975	Asn	2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC Phe	TIE	CAA Gln 985	AGT Ser	TTA Leu	AAA Lys .	Asp .	GAC Asp 990	CCA Pro	AGC Ser	2976
CAA Gln	AGC Ser	GCT Ala 995	AAC Asn	CTT Leu	TTA Leu	uta i	GAA (Glu) 1000	GCT . Ala	AAA Lys	AAG (Lys)	Leu i	AAT (Asn)	GAT (Asp)	GCT Ala	CAG Gln	3024
GCG Ala	CCG Pro 1010	Lys	GTA Val	GAC Asp	TAG											3042

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glú Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480

Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495

Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505

Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515

Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540

Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560

Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 575

Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590

Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 615 620

Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 635 640

Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
645 650 655

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670

Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685

Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700

Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720

Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
725 730 735

Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
740 745 750

- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 870 875 880
- Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895
- Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910
- His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925
- Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940
- Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960
- Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 965 970 975
- Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990
- Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005
- Ala Pro Lys Val Asp * 1010
- (2) INFORMATION FOR SEQ ID NO: 19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3509 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3509
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

1				5	A	911 F	ne A	sn T	yr A 10	\sn	Asp	Pro) Ile	e As 1	AT AAT Sp Asi .5	1
			20		J. J.	Lu F.	.0 2	25	ne A	та ,	Arg	Gly	Thi 30	Gl	G AGA y Arg	1
• "		35		,	, 5	4	10	sp A	rg I	ie 1	[rp	Ile 45	Ile	Pr	G GAA o Glu	
.	50		•	-, -,	5	5	.O G1	.u As	P P	ne A	sn 60	Lys	Ser	Se	GGT Gly	192
65			- 5	7	0	5 G1	u Iy	r ly	T AS	3p P 75	ro .	Asp	Tyr	Let	AAT Asn 80	240
ACT A		- .	,	5	••	C FII	e ne	9	n Tr 0	ır M	et :	Ile	Lys	Leu 95	Phe	288
AAT A Asn A	- J -	10	0	- <u>-</u>	S FI	ט דפי	10	y G1	u Ly	's L	eu I	Leu	Glu 110	Met	Ile	336
ATA A Ile A	11	iś		-	L Det	120) ARE) Arg	g Ar	g va	al P	25	Leu	Glu	Glu	384
	30	· ·			135	Val	. Ini	va.	L AS	n Ly 14	s L	eu :	Ile	Ser	Asn	432
CCA GO Pro Gl 145	-,			150))	nys	GIY	ile	155	≥ Al 5	.а А	sn I	Jeu	Ile	Ile 160	480
TTT GO		• ••	165	, , ,	Deu	VOII	GIU	170	GIT	ı Th	r I.	le A	sp :	[le [75	Gly	528
ATA CA Ile Gl		180)		261	Arg	185	GIÀ	Phe	: G1	y GI	ly I 1	le M 90	let	Gln	576
ATG AA Met Ly	19	5		914	-yr	200	SEI	vaı	rne	Ası	1 As 20	n V	al G	ln	Glu	624
AAC AA Asn Ly 21	0				215	VOII	Arg	Arg	GIA	220	Ph	e Se	er A	sp 1	Pro	672
GCC TTO Ala Lev 225			1100	230	GIU	Leu	TIE	HIS	235	Leu	. Hi	s Gl	ly L	eu 7	Tyr 240	720
GGC ATT	,-		245	vah	neu	PIO	TTE	250	Pro	Asn	Gli	u Ly	rs Ly 25	ys I 55	he	768
TTT ATO	G CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	116	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CT)	A TA 1 Ty 27	T Th	CA T	TT he	816

G GA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT	1536
GAT Asp	TTT Phe	TAA Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632

54	5			· A	59	. •	g As	Б ТТ	e Se	r Let 55!	u Th 5	ır Se	r Se	er P	he	Asp 560	1680
7.5	P A.	a Di	.u D	56	55	T AA r As	п цу:	s va.	1 Ty:	r Sei	r Ph	e Ph	e Se	r Me	et 75	Asp	1728
- 1		- <i></i> ,	58	10	.a As	T AA. n Ly:	s va.	585	GIV	ı Ala	a Gl	y Le	u Ph 59	e A] 0	la	Gly	1776
11	y va	5 9	5	.11 11	e va	A AA: l Asi	600	Phe	≀ Val	. Ile	: Gli	41. 60	a As	n Ly	's	Ser	1824
7.51	610)	C AS	p Ly	2 11	T GCA e Ala 615	Asp) lie	: Ser	Leu	620	e Vai	l Pro	э Ту	r :	lle	1872
625		· na	a De	u As	63	_	ASN	GIU	Inr	635	Lys	Gl _y	/ Ası	ı Ph	e (31u 340	1920
7.51	. UTC		e GI	64	5 5	A GGA a Gly	ATS	Ser	650	Leu	Leu	ı Glu	Phe	65!	e F 5	ro	1968
914	. Deu	. Der	66)	o val	A GTT L Val	GIĀ	A1a 665	Phe	Leu	Leu	Glu	Sex 670	Ту	r I	le	2016
nsp	VOII	675	, AS	ı nyı	, rie	ATT : Ile	680 LAS	Thr	Ile	Asp	Asn	Ala 685	Leu	Thi	L	ys	2064
A. 9	690	GIU	r pys	, iti	ser.	GAT Asp 695	Met	Tyr	GIÀ	Leu	Ile 700	Val	Ala	Gln	1 T:	rp	2112
705	502			. van	710	CAA Gln	Pne	Tyr	Thr	715	Lys	Glu	Gly	Met	T)	/T 20	2160
Lys	710	neu	ASI	725	GIN	GCA Ala	GIN	Ala	Leu 730	Glu	Glu	Ile	Ile	Lys 735	T	r	2208
Arg	TYL	VPII	740	lyr	ser	GAA Glu	rys	G1u 745	Lys	Ser /	Asn	Ile	Asn 750	Ile	As	p	2256
		755	116	ABII	361		760	ASN (GIU (GIA :	Ile	Asn 765	Gln	Ala	Il	.e	2304
p	770		uen	ASII	FIIE	ATA Ile 775	ABN (STA (Cys :	Ser \	Val 780	Ser	Tyr	Leu	Me	t	2352
785	nys	MEC	116	PIO	790	GCT (val (31u)	Lys 1	Leu I 795	Leu	Asp	Phe	Asp	As 80	n 0	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA /	AAT 1	yr :	ATA (Ile # BlO	SAT C	SAA . Slu .	AAT . Asn	AAA Lys	TTA Leu 815	TA Ty	T r	2448

TTC Leu	ATI	GG/	A AGT Ser 820		A GAA A Glu	TAI Tyr	GAA Glu	A AAA Lys 825	Ser	AAA Lys	GT#	AAI Asn	Lys 830	Туз	TTG Leu	2496
AAA Lys	ACC Thr	Ile 835	1100	Pro	TTT Phe	GAT Asp	CTI Leu 840	Ser	ATA	TAT	ACC Thr	AAT Asn 845	Asp	ACA Thr	ATA	2544
CTA Leu	ATA Ile 850	GIU	ATG Met	TTT Phe	AAT Asn	Lys 855	TAL	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	Leu	AAT Asn	AAT Asn	ATT	2592
ATC Ile 865	neu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	гÀЗ	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880	2640
TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GIU	GTA Val	TAT Tyr	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys	2688
AAT Asn	CAA Gln	TTT	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr	2736
CAA Gln	AAT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Lėu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val	2784
AGC Ser	TTT Phe 930	TGG Trp	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn	2832
TAT Tyr 945	ATT Ile	CAT His	AAT Asn	GAA Glu	TAT Tyr 950	ACA Thr	ATA Ile	ATT Ile	AAT Asn	TGT Cys 955	ATG Met	AAA Lys	AAT Asn	AAT Asn	TCG Ser 960	2880
GGC Gly	TGG Trp	AAA Lys	ATA Ile	TCT Ser 965	ATT Ile	AGG Arg	GGT Gly	AAT Asn	AGG Arg 970	ATA Ile	ATA Ile	TGG Trp	ACT Thr	TTA Leu 975	ATT Ile	2 92 8
GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	TCG Ser	GTA Val 985	TTT Phe	TTT Phe	GAA Glu	Tyr	AAC Asn 990	ATA Ile	AGA Arg	2976
GAA Glu	GAT Asp	ATA Ile 995	TCA Ser	GAG Glu	TAT Tyr	TTE	AAT Asn 1000	Arg	TGG Trp	TTT Phe	TTT Phe	GTA Val 1005	Thr,	ATT Ile	ACT Thr	3024
ABR .	AAT Asn 1010	Leu	AAT Asn	AAC Asn	GCT . Ala	AAA Lys 1015	ATT Ile	TAT Tyr	ATT Ile	Asn	GGT Gly 1020	Lys	CTA Leu	GAA Glu	TCA Ser	3072
AAT Asn 1025	Inr	GAT Asp	ATT Ile	rys	GAT Asp 1030	ATA . Ile .	AGA Arg	GAA Glu	Val	ATT Ile 1035	Ala	AAT (Asn (GGT Gly	Glu	ATA Ile 1040	3120
ATA Ile	TTT Phe	AAA Lys	ren .	GAT Asp 1045	GIA '	GAT . Asp	ATA (Asp .	AGA Arg 1050	ACA Thr	CAA Gln	TTT . Phe	Ile	TGG Trp 1055	ATG Met	3168
AAA '	TAT Tyr	Pne	AGT Ser 1060	He	TTT : Phe :	AAT . Asn '	Thr (GAA Glu 1065	TTA . Leu	AGT (CAA Gln	Ser	AAT Asn 1070	ATT	GAA Glu	3216
GAA /	Arg	TAT Tyr 1075	AAA Lys	ATT :	CAA '	Ser '	TAT Tyr 1080	Ser	GAA Glu	TAT '	Leu	AAA (Lys) 1085	GAT Asp	TTT Phe	TGG Trp	3264

GGA Gly	AAT Asn 109		TTA Leu	ATG Met	TAC	AAT Asn 109	Lys	GAA Glu	TAT	TAT Tyr	ATO Met	Phe	AAT Asn	GCG Ala	GGG Gly	3312
110	5	7.571	501		1110) rya	reu	гÀг	Lys	Asp 111	Ser 5	Pro	Val	Gly	GAA Glu 1120	3360
ATT Ile	TTA Leu	ACA Thr	CGT Arg	AGC Ser 1125	AAA Lys	TAT Tyr	AAT Asn	CAA Gln	AAT Asn 113	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1135	Tyr	3408
AGA Arg	GAT Asp	TTA Leu	TAT Tyr 1140	116	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1145	He	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	3456
TCT Ser	CAA Gln	TCT Ser 1155	110	AAT Asn	GAT Asp	ASP	ATA Ile 1160	vai	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3504
CTA Leu	GA															3509

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1169 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
35 40 45

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

PCT/GB97/02273

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 710 Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 855

Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly

870

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met
1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

		G CO t Pr 1	CA (GTT Val	ACI Thi	- 11	A AA e As 5	T AA n As	T TT n Ph	T AA e As	T TA' n Ty:	r Ası	T GA n As	T CC P Pr	T AT	e As	AT AA Sp As .5	AT in	48
	****				20)	c Me	c GI	u PI	2	-	A A L	a Ar	g Gl	y Th 3	r Gl O	y Ar	g	96
	-7.	7		35	710		z Ly:	3 11(2 11.	O AS	r CG1	, Ile	Tr	0 Il	e Il 5	e Pr	o Gl	u	144
	••••	5	ō	****	F 11.C	GI	. 171	. Ly:	FIC) GI	GAT JAS	Phe	Asr 60	l Ly:	s Se	r Se	r Gl	Y	192 ·
	65	5	.	.511	AL Y	vəb	70	. Cys	GIL	т тух	TAT Tyr	Asp 75	Pro) Asp	ту1	Le	1 Ası 18	n)	240
	****	, na		.sp	пуs	85 85	ASI	TIE	Pne	. Leu	CAA Gln 90	Thr	Met	Ile	: Lys	Let 95	ı Phe	•	288
			, -	-6	100	ser	LIYE	PIO	Leu	105		Lys	Leu	Leu	Glu 110	Met	Ile	:	336
	*10		1:	15	116	FIO	TYL	rea	120	Asp	AGA Arg	Arg	Val	Pro 125	Leu	Glu	Glu	l	384
		130)		usii	116	NIG	135	vai	Int	GTT Val	Asn	Lys 140	Leu	Ile	Ser	Asn		432
	145	O.	G 2	, ,	val	GIU	150	Lys	Lys	GIÀ	ATT Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160		480
	• •••	GIY		.0 (31 y	165	Val	reu	Asn	GIU	AAT Asn 170	Glu	Thr	Ile	Asp	Il e 175	Gly		528
	116	J 111	AQ	1	.80	FIIC	WIG	ser	Arg	185	Gly	Phe	Gly	Gly	Ile 190	Met	Gln		576
•	7G L	пyъ	19	5	ys .	PIO	GIU	TYE	200	ser	GTA Val	Phe .	Asn	Asn 205	Val	Gln	Glu		624
•	1911	210	GI	уд	ıta i	ser	11e	215	Asn	Arg	CGT (Gly :	Tyr 220	Phe	Ser	Asp	Pro		672
•	ICC la 25	TTG Leu	AT:	A T e L	TA A	יופנ	CAT (His (230	GAA Glu	CTT Leu	ATA Ile	CAT (GTT : Val 1 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240		720

GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245) Asp	TTA Leu	CCA Pro	ATI	GTA Val 250	Pro	AAI Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT Phe	Met	CAA Gln	TCT Ser 260	Inr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	816
GGA Gly	GGA Gly	CAA Gln 275	Asp	Pro	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC	864
TAT Tyr	GAT Asp 290	гÀг	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
TAA Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	TAA nsA	GAA Glu	GAT Asp	Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	TAA neA	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

ASI	o Pne	519	n val	L As	o va.	Pro	520	l Tyr	r Glu	ı Ly:	s Gl	1 Pro 529	o Al	a Il	A AAA e Lys	1584
AAA Lys	A ATT	Phe	Γ ACA ∃ Thi	GAT Asp	GAZ Glu	AAT Asn 535	Thi	C ATO	TTT Phe	CAJ Glr	TA: 1 Ty: 540	: Let	TA Ty	C TC r Se	T CAG r Gln	1632
ACA Thr 545	Phe	CCI Pro	CTA Leu	GAT Asp	11e 550	: Arg	GAT Asp	T ATA	AGT Ser	TTA Lev 555	Thi	TCI Ser	TC:	A TT	GAT SASP S60	1680
GAT Asp	GCA Ala	TTA Leu	TTA Leu	TTI Phe 565	Ser	AAC Asn	AAA Lys	GTT Val	TAT Tyr 570	Sex	TTT	TTI Phe	TC: Sei	ATC Mei 579	G GAT Asp	1728
Tyr	Ile	Lys	580	Ala	Asn	Lys	Val	Val 585	Glu	Ala	Gly	Leu	9he	Ala	A GGT A Gly	1776
Trp	val	Lys 595	Gin	He	Val	Asn	Asp 600	Phe	Val	Ile	Glu	Ala 605	Asn	Lys	AGC Ser	1824
Asn	610	Met	Asp	Lys	He	Ala 615	Asp	Ile	Ser	Leu	11e 620	Val	Pro	Туг	ATA Ile	1872
GGA Gly 625	Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTI Phe	GAA Glu 640	1920
Asn	Ala	Phe	Glu	11e 645	Ala	Gly	Ala	Ser	Ile 650	Leu	Leu	Glu	Phe	Ile 655	CCA Pro	1968
GIU	Leu	Leu	660	Pro	Val	Val	Gly	Ala 665	Phe	Leu	Leu	Glu	Ser 670	Tyr		2016
Asp	Asn	675	Asn	Lys	Ile	Ile	Lys 680	Thr	Ile	Asp	Asn	Ala 685	Leu	Thr	_	2064
Arg	690	Glu	Lys	Trp	Ser	GAT Asp 695	Met	Tyr	Gly	Leu	Ile 700	Val	Ala	Gln	Trp	2112
705	ser	Thr	val	Asn	710	CAA Gln	Phe	Tyr	Thr	Tle 715	Lys	Glu	Gly	Met	Tyr 720	2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	2208
AGA Arg	TAT Tyr	TAA Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA . Glu .	Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	Ile	AAC Asn 750	ATC Ile	GAT Asp	2256
TTT .	AAT Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys :	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	2304
GAT . Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	Phe	ATA I Ile I 775	AAT Asn	GGA Gly	TGT Cys	Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	2352

AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 858 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

- Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205
- Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 210 215 220
- Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 235 230 235
- Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 250 255
- Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 265 270
- Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 275 280 285
- Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 290 295 300
- Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 315 320
- Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325
- Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 340 345 350
- Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 355 360 365
- Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370
- Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 385 390 395 400
- Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 410 415
- Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 420 425 430
- Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435
- Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 450 455 460
- Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 470 475 480
- Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 495
- Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 500 505 510
- Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 515 520 525
- Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 530

- 100 -

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 600 Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

WO 98/07864 PCT/GB97/02273

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

rie i	G CC t Pro	A GT O Va	T ACA	A ATA	A AAT E Asr	TAAT Asn	TTT	C AA1 e Asr	TAT Tyr 10	Asr	GAT Asp	CCT Pro	T AT	T GA' e As _i	T AAT D Asn	48
AA? Asi	AA? ASI	T AT	T ATT	: met	ATC Met	GAG Glu	Pro	CCA Pro 25	> Phe	GCG Ala	AGA Arg	GGT Gly	Thi	r Gly	G AGA / Arg	96
TAT Tyr	TA1	AAI Lys	s MT9	TTI Phe	AAA Lys	ATC	ACA Thr 40	Asp	CGT Arg	ATT	TGG	ATA Ile 45	Ile	CCC Pro	G GAA	144
AGA Arg	TAT Tyr 50	TUI	TTT Phe	GGA Gly	TAT	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGT Ser	TCC Ser	GGT	192
ATT Ile 65	Pne	AAT Asr	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	240
1111	ASII	Asp	гуз	85 Lys	ASI	IIe	Pne	Leu	Gln 90	Thr	Met	Ile	Lys	Leu 95	TTT	288
ASII	Arg	me	100	ser	гЛа	Pro	Leu	105	Glu	Lys	Leu	Leu	Glu 110	Met		336
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
rne	130	inr	AAC Asn	Tie	Ата	135	Val	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn	432
145	GIÀ	GIU	GTG Val	Glu	150	Lys .	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160	480
Pne	GIY	Pro	GGG Gly	165	vai	Leu	Asn	Glu	Asn 170	Glu	Thr	Ile	Asp	Ile 175	Gly	528
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA . Ser .	Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	Gly	ATA Ile 190	ATG Met	CAA Gln	576
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	Asn .	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT . Ser	iie .	TTT Phe 2	AAT . Asn .	AGA Arg	CGT (Arg (Gly '	TAT Tyr 220	TTT ' Phe :	TCA Ser	GAT Asp	CCA Pro	672

GCC Ala 225	. דבת	ATA Ile	TTA Leu	AIG Met	CAT His 230	GIU	CTI Leu	ATA Ile	CAT His	GT1 Val 235	. Leu	CAT His	GG#	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245	Asp	TTA Leu	CCA	ATT Ile	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT Phe	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
ser	Asp	Lys	Asp	Met 405	GAA Glu	Lys	Glu	Tyr	Arg 410	Gly	Gln	Asn	Lys	Ala 415	Ile	1248
TAA neA	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488

- 103 -

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1632
ACA Thr 545	TTT Phe															1644

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu
35 40 45

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu

- 104 -

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 245 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 410 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 520 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln 535 Thr Phe Pro Leu 545

- 105 -

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2616 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATG Met	GI	TTO	C GTO	G AA(L Asr	AAC Lys	G CAG	TTC Phe	AAC Asr	TAT Tyr	Lys	GAC Asp	CC1	GTA Val	AA(Asi	GGT Gly		48
			20)		. Lys	116	25	Asn	Ala	Gly	Glr	Met 30	Glr	CCG Pro		96
	2 , 2	35		. nys	116	nis	40	гåз	ile	Trp	Val	Ile 45	Pro	Glu	CGC Arg	1	144
7.02	50			nsii	PIO	55	Gru	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	GAA Glu	1	.92
65	Бys	GIII	vai	PIO	70	TCA Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80	2	40
AOP	nan	G L u	Буз	85	ASII	TAC Tyr	Leu	rys	90	Val	Thr	Lys	Leu	Phe 95	Glu	2	88
.,29		-7-	100	1111	vab	CTG Leu	GIY	105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	3.	36
	U.J	115	110	FILE	Пр	GGT Gly	120	ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys	31	84
V 44.1	130	vaħ	1111	ASII	Cys	ATT Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr	43	32
145	SCI	GIU	Gru	nea	150	CTC Leu	vaı	lle	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	48	30
ATC (9211	rne	GIU	165	rys	ser	rne (GIY	H18 170	Glu	Val	Leu	Asn	Leu 175	Thr	52	28
CGT A	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT (GID :	TAC Tyr 185	ATT Ile .	CGT '	TTC Phe	Ser	CCA (Pro 1	GAC Asp	TTC Phe	57	6

ACG Thr	TTC Phe	GGT Gly 195	FILE	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GIU	GTT Val	GAT Asp	ACC Thr	AAC Asr 205	Pro	CTC Lev	TTG Leu	624
GGT Gly	GCA Ala 210	GIY	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	Leu	GCA Ala	CAC His	GAG Glu	672
225	116	urs	Ala	GIY	230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	AAC Asn 240	720
Arg	vai	rne	гÀа	245	ASN	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255		768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	A AA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA . Leu .	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT . Ile .	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT .	AAA ' Lys :	TCA Ser	TTA Leu	GAT Asp	Lys	GGA Gly 445	TAC Tyr	TAA naA	AAG Lys	1344
Ala	TTA Leu 450	TAA naA	GAT Asp	TTA Leu	Cys	ATC . Ile : 455	AAA (Lys '	GTT Val	AAT Asn	Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

465	5			GAT Asp	470	****		1511	Asp	475	Asn	Lys	Gl	y Gl	lu (31u 180	1440
				ACT Thr 485		116 0	ilu r	ııa .	490	Glu	Glu	Asn	Ile	9 S€	r I	Leu	1488
			500	CAA Gln	-y	LYL L	5	05	rne	Asn	Phe	Asp	Asr 510	ı Gl	u P	ro	1536
		515	501	ATA (,	5	20	er s	er	Asp	Ile	Ile 525	Gly	Gl	n L	eu	1584
	530			AAT A Asn 1	5	35	ra P	ue 1	TO.	Asn	Gly 540	Lys	Lys	Ty:	r G	lu	1632
545		-,-	- , -	_	50	ne n	is i	YF L	eu :	555	Ala	Gln	Glu	Phe	G. 56	lu 50	1680
	7	-,,,		AGG A Arg I 565	TC A	ra De	au II	1F A	3n S	ser '	Val	Asn	Glu	Ala 575	Le	u	1728
_			580	CGT G Arg V	 -;	y 1 - 11:	58	15 P	ne s	ser s	ser .	Asp '	Tyr 590	Val	Ly	's	1776
-2-		595	_,,,,,,	GCT A	G.	60	0 0	a me	פל פ	he I	en (Gly 7 605	Lrp	Val	Gl	u	1824
(610		-,	SAT T	61	5	b GI	u II	ır s	er G	20	Val S	Ser	Thr	Th	r	1872
625	-,-			AT AI 11 qe 63	0	T 11	= 116	e 11	.e P	ro T 35	yr 1	(le G	ly	Pro	A1:	aa O	1920
			6	AT AT sn Me 45	C DC	u iyi	гъ	65	P A	sp P	he V	al G	ly	Ala 555	Le	1	1968
		6	60	CT GT la Va		e net	665	GI	u Pr	ne I.	le P	ro G 6	lu 1 70	lle	Ala	ı	2016
ATA C	6	75	.	~y	r 511/	680		va.	ı Se	r T	r I	le A: 85	la A	sn	Lys	1	2064
	90		41 G.		695	Asp	ABN	Ala	a Le	u S∈ 70	Er L	ys A:	rg A	sn	Glu		2112
AAA TO Lys To 705		- p	'	710	. <i>Dy E</i>	TYL	116	va.	71	r As 5	n T	rp Le	eu A	la	Lys 720		2160
GTT A	AT A	CA CA hr Gl	AG A1 in I1 72	e Ast	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	: Ly	A AT s Me	G AJ	AA GA /s Gl	u A	CT 1 la 1 35	ITA Leu		2208

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
					ATT Ile 870	AAG Lys	TAA *									2616

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

- 109 -

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435 440 445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 550 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 650 Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 680 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805 810

WO 98/07864 PCT/GB97/02273

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855 860

Thr Phe Thr Glu Tyr Ile Lys 865 870

- (2) INFORMATION FOR SEQ ID NO: 27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA CCATCAACAA CTTCAACTAC AACGACCCGA TCGACAACAA CAACATCATC 60 ATGATGGAAC CGCCGTTCGC ACGTGGTACC GGTCGTTACT ACAAGGCTTT CAAGATCACC 120 GACCGTATCT GGATCATCCC GGAACGTTAC ACCTTCGGTT ACAAACCTGA GGACTTCAAC 180 AAGAGTAGCG GGATTTTCAA TCGTGACGTC TGCGAGTACT ATGATCCAGA TTATCTGAAT 240 ACCAACGATA AGAAGAACAT ATTCCTTCAG ACTATGATCA AGTTATTTAA TAGAATCAAA 300 TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA 360 GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA 420 TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA 480 TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT 540 TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA 600 AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT 660 TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTTATAC ATGTTTTACA TGGATTATAT 720 GGCATTAAAG TAGATGATTT ACCAATTGTA CCAAATGAAA AAAAATTTTT TATGCAATCT 780 ACAGATGCTA TACAGGCAGA AGAACTATAT ACATTTGGAG GACAAGATCC CAGCATCATA 840 ACTCCTTCTA CGGATAAAAG TATCTATGAT AAAGTTTTGC AAAATTTTAG AGGGATAGTT 900 GATAGACTTA ACAAGGTTTT AGTTTGCATA TCAGATCCTA ACATTAATAT TAATATATAT 960 AAAAATAAAT TTAAAGATAA ATATAAATTC GTTGAAGATT CTGAGGGAAA ATATAGTATA 1020 GATGTAGAAA GTTTTGATAA ATTATATAAA AGCTTAATGT TTGGTTTTAC AGAAACTAAT 1080 ATAGCAGAAA ATTATAAAAT AAAAACTAGA GCTTCTTATT TTAGTGATTC CTTACCACCA 1140 GTAAAAATAA AAAATTTATT AGATAATGAA ATCTATACTA TAGAGGAAGG GTTTAATATA 1200

TCTGATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATGAAGAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGTTTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAGAATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAAAAAC	AACCCGCTAT	AAAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТАААААТАА	AATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
aatgagggta	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
CAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2574 base pairs
 (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

ATGCCAGTTA	CAATAAATAA	TTTTAATTAT	AATGATCCTA	TTGATAATAA	TAATATTATT	60
ATGATGGAGC	CTCCATTTGC	GAGAGGTACG	GGGAGATATT	ATAAAGCTTT	TAAAATCACA	120
GATCGTATTT	GGATAATACC	GGAAAGATAT	ACTTTTGGAT	ATAAACCTGA	GGATTTTAAT	180
AAAAGTTCCG	GTATTTTTAA	TAGAGATGTT	TGTGAATATT	ATGATCCAGA	TTACTTAAAT	240

ALAGARIAI AIIITTACAA ACAATGATCA AGTTATTTAA TAGAATCAAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	480
TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	540
TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCCTCCA CGGTCTCTAC	720
GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	780
ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCCG TGGTATCGTC	900
GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	1020
GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1200
AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1260
TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1320
AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAAA	1500
ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
TACGAAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	1740
GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	2160
AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

WO 98/07864 PCT/GB97/02273

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	234
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	240
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
ICAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

CLAIMS

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain $H_{\rm N}$ portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
- 8. A polypeptid according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated H_{N} of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polyp ptide according to Claim 14 wherein on or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin h avy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated $H_{\rm c}$ of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said s cond polypeptide is

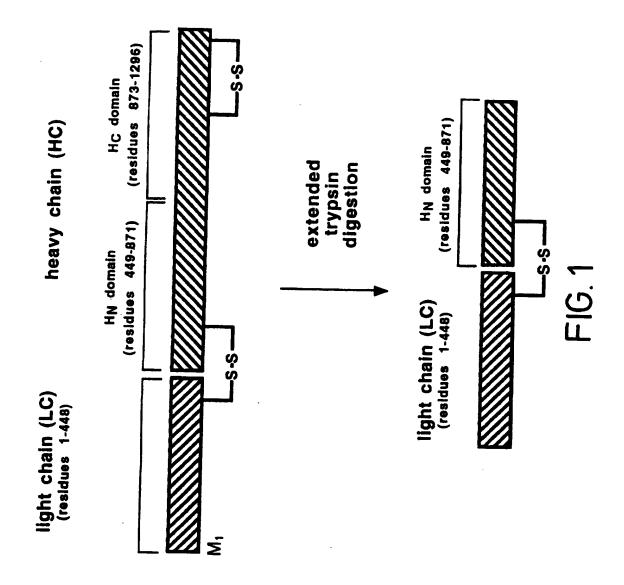
adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

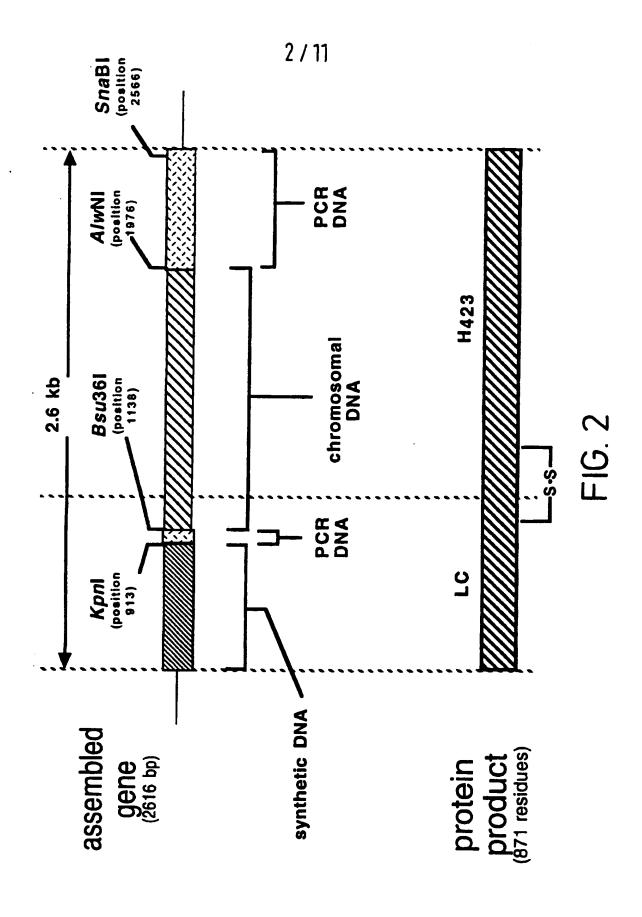
- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

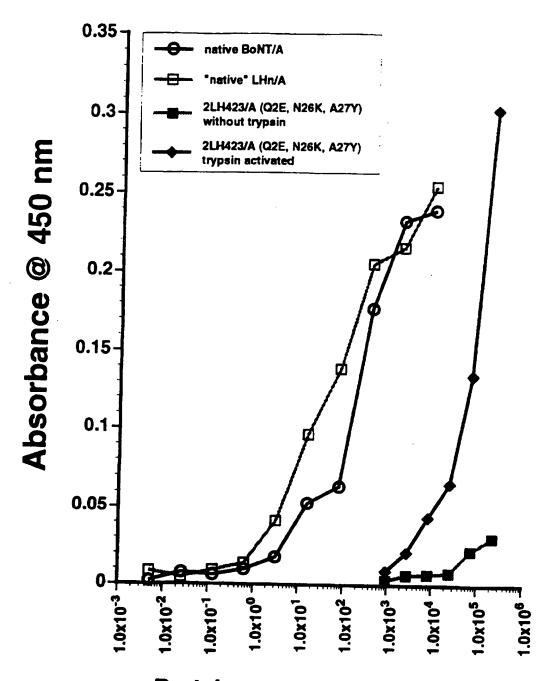
- 40 and comprising nucl otides encoding residues 1-448 of a botulinum toxin type A light chain.
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain $H_{\rm N}$ domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.







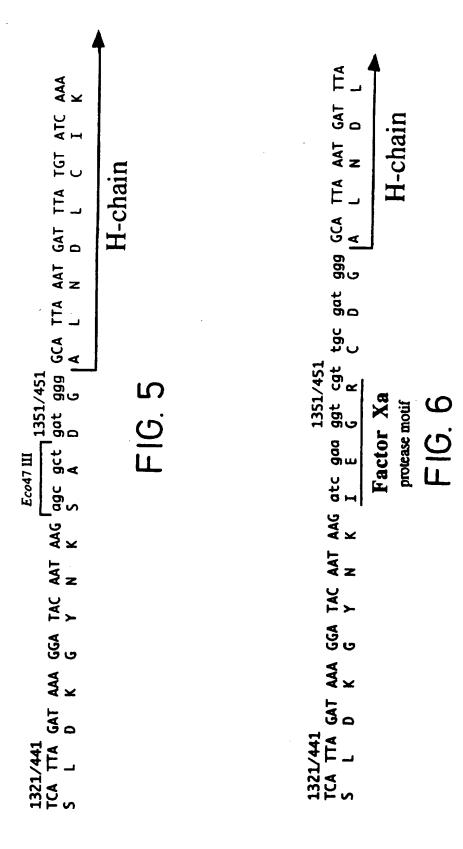
Protein concentration (ng/ml)

FIG. 3

			4/	11		
7		₹	9			
DIAYIKIPNAGQMQPV (Seq 1.D. 2)		I.D. 4)	p v (Seq I.D. 6)			
- b		-	<u> </u>			
S)		P V (Seq	(Se			
>		>	>	>	>	
۵		٩	۵	٩	۵	
0		0	0	0	0	
Σ		T	Σ	Œ	Σ	
0		0	0	0	0	
9		ی	<u> </u>	<u> </u>		
4		>	>	⋖ -	4	
اچا		-	<u>~</u>			
-		H	H	ПР	Н	
 ⊻		×	×	×	×	
H		H	H	H	H	
>		>	>	>	>	
⋖		⋖	⋖	⋖	⋖	
H	<u> </u>	H	H	H	H	
۵	ب	0	a	0	0	
>	ш	>	>	>	>	
9 Z	u_	9	9	G	ပ	
	m m	Z	Z	Z	Z	
>		>	>	>	>	
0 D	Q K	0 P	0 P	9 0	0 P	
×	-	~	×	×	¥	
>		<u></u> ⊢	-	>	→	
z	~	z	z	Z	Z	
u.	-	L	ш.	щ	ı <u>ı</u>	
0	S	0	0	0		
¥	-	¥	×	¥	×	
Z	X	Z	Z	Z	Z	
>	=	>	>	>	>	
_	H		ш.	<u> </u>		
MQFVNKQ	GSPGIHMTS	G T M E F V N K Q	GSMEFVNKQ 123	1 Q F V N K Q	MPFVNKQ	
2 7		T M 24	- M	3 H	≥ ⊣	
	. છ ન	_ 	2.2			
			•			Γ
		_	_	8		L
		(Q ₂ E, N ₂₆ K, A ₂₇ Y)	2LH423/A (Q2E, N26K, A27Y)	Native BoNT/A, C. botulinum 2169 Thompson et al.1990	⋖	
		₹	Š.	4 2 B	₹,80 8	
ĕ	§	26K	'A 26K	Native BoNT/A, C. botulinum 2169 Thompson et al.19	Native BoNT/A, C. botulinum 62A Binz et al.1990	
LH ₄₂₃ /A	1423	Z ui	423. N	ulin USOI	Bo ulin t at.	
≐	23LH423/A	S.	2LH423/A (Q2E, N26	tive bot	tive bot	
	ผ	=	~ C	R C F	Native BoNT/A, C. botulinum 62 Binz et al.1990	

F1G 4

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.



6/11

ACA T CAG GAC D TAT Y GGA G ATG M CCT P 6CG A GAG E 16T C GTG V AGG R CTG L AGT CGG S R AGG R 11C F CAG AGG R AGC S , 5TG > E AA TAT Y AGC S ပ္ပ်ပ္ပ ပ ر 30 GAG E F F ACA T TCA S GCT A ر ا ا AAG K 767 C 999 AAG K 787 **≸**~ 76C C AAC N GAG F CCT P 77 T F AAG K GAT D TAT Y GTG V 2587/863 TAC GTA Y V 2647/883 CCG GAG P E 2707/903 GGC TTT G F 2767/923 GGT ATC G I 2827/943 GCA CCC

=16.7

H CAT ACA T TCT S AAG K GAT D ATT ATT I TAT Y GAT D GAA E 2617/873 TTT ACT F T 2677/893 TAT CAA Y Q ر و و ACA T TCT S TCA S H 4 ¥ L L ATA I ¥ L AGA R **₹**~ AGA R **₹**⊘ ¥ × E SA A A A AAG GAT K D GAT D Ę,

F16.8

CAA CCA A GCC AAC N **\$** & A A GAA E GAA E TTA L AAA K CGA R GAA E AAA X GAA E CAA O AAC N EJ_ AAG K AAC N AAC GAA E GCT A A A A GAA E TAT Y AAC N AGC S AAC N AAC N 2617/873
TTT ACT GAA
F T E
2677/893
GAT GAA GCC
D E A
2737/913
CAT TTA CCT
H L P
2797/933
CCA AGC CAA
P S Q
2857/953
AAA GTA GAC
K V D SAC H GAC D 6CG A GAT D RAT H TTA L 900 8 GAG K K K CAG TTA L GAC ¥ L TAT Y |≺ **¥** GCT A GAT CAA AGT Ę GAT D GAG E A SCT **§** ~ 900 A AAT N TAT Y ¥ z ATC I CTA

8/11 LH₄₂₃/A

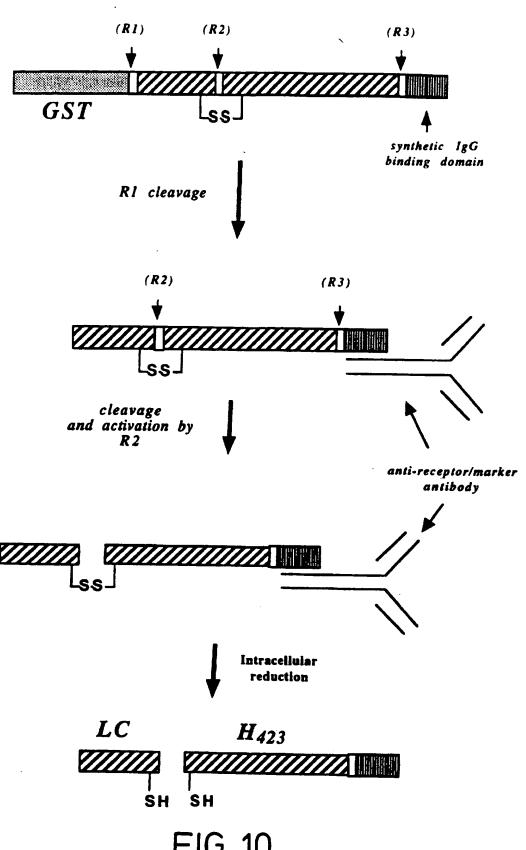
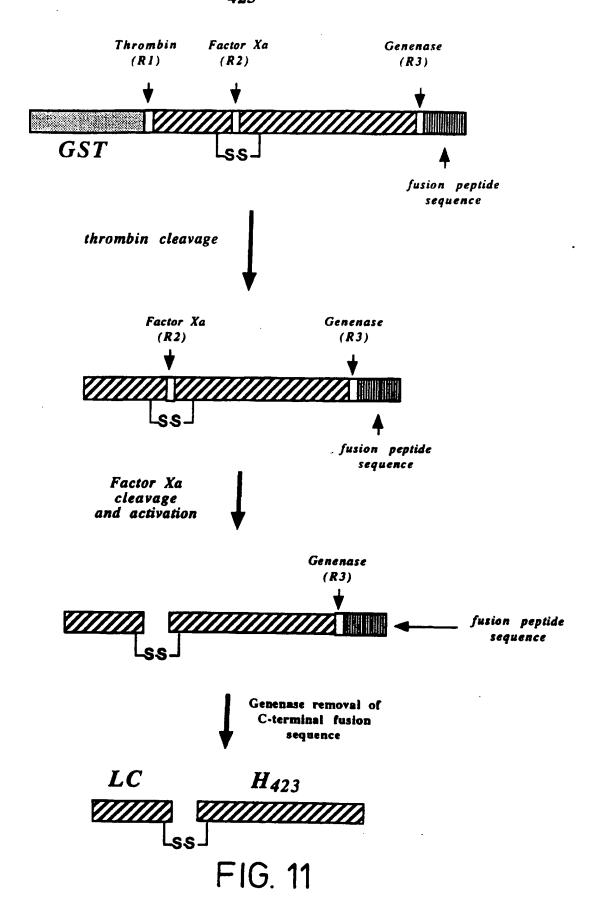


FIG. 10

$LH_{423}/A^{9/11}$



10/11

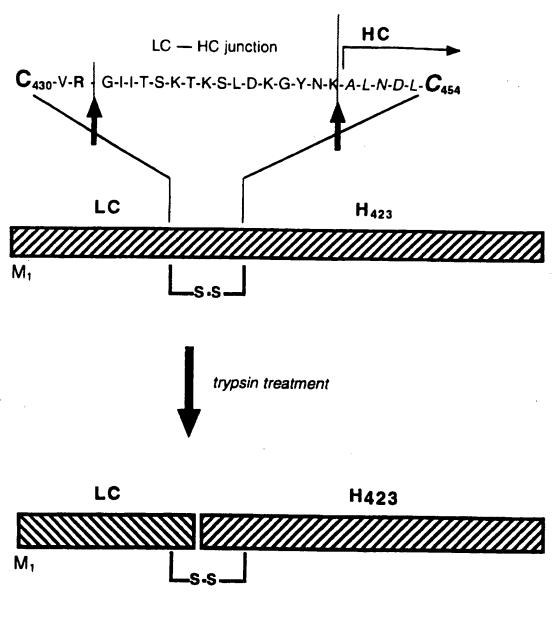
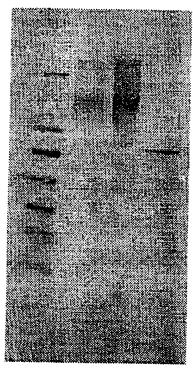


FIG. 12

11 / 11

Panel A. 1 2 3 4



Panel B. 1 2 3 4

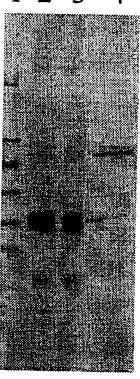


FIG. 13

INTERNATIONAL SEARCH REPORT

inten unal Application No

PCT/GB 97/02273 CLASSIFICATION OF SUBJECT MATTER PC 6 C12N15/31 C12N IPC 6 C12N1/21 C07K14/33 C12P21/02 A61K38/16 A61K39/08 According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P A61K Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. X WO 96 12802 A (OPHIDIAN PHARM INC 1-52 ;WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document X KURAZONO H ET AL: "Minimal essential 1-52 *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II -/--Further documents are listed in the continuation of box C. X Patent family members are listed in annex. IX I Special categories of cited documents : *T* later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not considered to be of particular relevance cited to understand the principle or theory underlying the invention "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) involve an inventive step when the document is taken alone "Y" document of pertioular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu-"O" document referring to an oral disclosure, use, exhibition or other means ments, such combination being obvious to a person skilled *P* document published prior to the international filing date but later than the priority date claimed "&" document member of the same patent family Date of the actual completion of the international search Date of mailing of the international search report 3 0. OL 98 9 December 1997 Name and mailing address of the ISA **Authorized officer** European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni,

Hillenbrand, G

Fax: (+31-70) 340-3016

INTERNATIONAL SEARCH REPORT

Interconal Application No
PCT/GB 97/02273

		PCT/GB 97	/02273
C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.
X .	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document		1
-	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document		1,26,35
		·	
		•	

INTERNATIONAL SEARCH REPORT

tnter. .onal Application No
PCT/GB 97/02273

Patent document	Publication	D.AA.	
cited in search report	date	Patent family member(s)	Publication date
WO 9612802 A	02-05-96	US 5601823	A 11-02-97
		US 5196193 /	A 23-03-93
		AU 3968395 /	A 15-05-96
		EP 0796326 /	A 24-09-97
		FI 971732 /	A 23-06-97
		NO 971868 A	
		PL 320214 A	
		ZA 9508990 A	
		AU 6653894 A	04-07-94
		CA 2150935 A	23-06-94
		EP 0671902 A	20-09-95
		WO 9413264 A	
		US 5466672 A	
		US 5599539 A	
		AU 638786 B	
		AU 6895191 A	
		EP 0498854 A	
		WO 9106306 A	
		US 5443976 A	
		US 5340923 A	